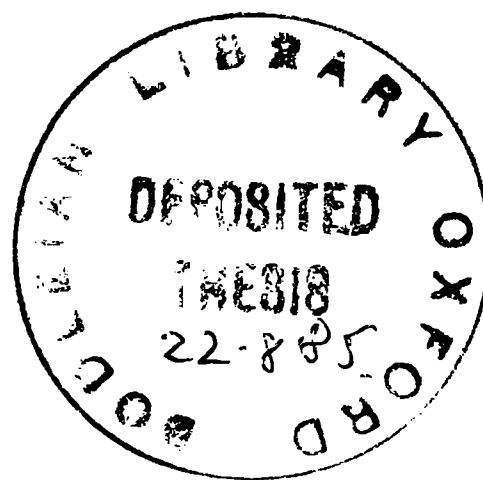


A U T O N O M Y A N D P A T E R N A L I S M

by

Yolande Westwell-Roper

Linacre College, Oxford



Submitted to the Sub-Faculty of Philosophy

for the degree of
Doctor of Philosophy

University of Oxford
Hilary Term, 1985

A B S T R A C T

AUTONOMY AND PATERNALISM

Yolande Westwell-Roper, Linacre College, Oxford.

Submitted for the degree of D. Phil., Hilary Term, 1985.

The aim of the thesis is to develop a theory of autonomy-respecting paternalism which delineates an area of morally permissible paternalistic interference. Paternalism is defined as any infringement of a prima facie right to non-interference for the purpose of protecting or promoting the recipient's well-being. It is argued that autonomy involves not only liberty of action, but also the achievement of self-construction and self-control. The role of rational reflection in the achievement of self-construction is discussed at some length, and the importance of knowledge, including self-knowledge and moral sensitivity is emphasized. The right to non-interference is taken to be grounded in the intrinsic value of autonomy, and possession of a prima facie right to non-interference is extended to all beings with desires. This general right is discussed in terms of three rights of greater specificity: non-interference with actions, states, and opportunities. An account of the vehicle for alienation of the right to non-interference is developed in terms of a technical notion of subsequent approval. The principle of respect for autonomy is shown to be as applicable in paternalistic dealings with children as it is with adults, without this having counter-intuitive consequences in practice. It is also shown how far the paternalistic promotion of a recipient's well-being, understood as the satisfaction of informed desires, can be reconciled with the principle of respect for autonomy. Finally, the theory is applied to particular cases of paternalism in familial, medical, and legal contexts.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	v
INTRODUCTION	1
A. Purpose of the Study	1
B. Definition of Paternalism	6
C. The Importance of Autonomy	11
D. Justification	16
E. Overview	20
Chapter	
I AUTONOMY	24
A. Full Autonomy	29
1. Liberty	29
2. Self-Construction	30
(a) Rational Reflection	33
i. Stability and Continuity	33
ii. Control	39
iii. Components	44
(b) Strong Evaluation	51
3. Self-Control	54
(a) Governing the Peripheral Self	54
(b) Failures of Self-Government	58
(c) Self-Creation	64
B. Objections	67
1. Spontaneity	67
2. Autonomy and the Whole of Human Freedom	72
3. The Argument from Socialization	74
II THE RIGHT TO NON-INTERFERENCE	78
A. Source	79
B. Possession	81
C. Infringements	95
D. Alienation	101
1. Conditions	101
2. Objections	111
(a) Adaptive Preference Formation	111
(b) The Reductionist View of Personal Identity	113

TABLE OF CONTENTS Cont'd.

Chapter		
III	WELL-BEING	117
	A. Alternative Accounts of Well-Being	118
	1. Objective List Theories	118
	2. The Mental State Account	121
	3. Well-Being as Desire Satisfaction	122
	B. Implications	130
	1. Autonomy and Well-Being	130
	2. Basic Goods	134
	3. Well-Being and Perfectionism	139
IV	PATERNALISM	145
	A. General Implications	146
	1. Incompetence and Paternalism	146
	(a) The Relevance of Incompetence	146
	(b) Incompetence and Self-Regarding Decisions	156
	i. Ignorance	157
	ii. Compulsion	159
	iii. Non-Rationality	162
	iv. Irrationality	164
	(c) Children	169
	2. Consent and Paternalism	170
	B. Practical Implications	172
	1. Inter-Personal Paternalism	173
	2. Legal Paternalism	191
	3. Education	204
	C. Conclusion	208
	NOTES	212
	BIBLIOGRAPHY	215

A C K N O W L E D G E M E N T S

I am very grateful to my supervisor, Jonathan Glover, for his many helpful comments and criticisms during the preparation of this thesis. I am also very grateful to Professor Richard Hare and the late John Mackie for their helpful comments and supervision at earlier stages.

I N T R O D U C T I O N

A. Purpose of the Study

The purpose of this study is to present a theory to enable us to distinguish between cases of justified and unjustified paternalism. Paternalistic interventions in which it seems clear we are fully justified in acting against someone's wishes for his own good are common occurrences. We do not, for example, permit young children to play with matches or sharp knives, regardless of their wishes. There are also many cases in which paternalistic intervention seems obviously wrong. We would not, in ordinary circumstances, take a fully competent adult's cigarettes away from him, even if we are certain that not smoking would be better for him. But there are also many cases of paternalism in which the question of justification seems puzzling and controversial. It is often unclear whether doctors are justified in giving blood transfusions, for example, and other kinds of treatments to unwilling members of certain religious groups. It is equally unclear under what conditions, if any, a family is justified in lying to one of its members about the terminal nature of that member's illness. Similarly, practices such as forcing mentally competent elderly persons (judged to be at risk) to leave their homes and enter state institutions, are also morally controversial. An adequate theory of the justification of paternalism will help us decide what ought

to be done in these and other related cases.

It is generally agreed that preventing harm to others is often a sufficient warrant for interfering with someone, but there is little agreement about when a person may be "compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinion of others, to do so would be wise or even right" (Mill, On Liberty, p.13). It is the task of this study to clarify the factors which provide sufficient warrant for paternalistic intervention. In certain cases such factors may be morally outweighed from an all-things-considered point of view by considerations of preventing harm to others, but that contingency is not of central importance here. This study is about the justification of paternalism qua paternalism.

There are strong and weak versions of the justification of paternalism. Strong paternalism is the view that paternalistic interference with the informed, voluntary wishes of competent persons is sometimes permissible regardless of the recipient's own conception of the good. Plato's views in The Republic are an obvious example of strong paternalism. Weak paternalism is the view that some defect of will such as irrationality, ignorance, or some other form of involuntariness on the part of the recipient is necessary for the justification of paternalism. Mill is ordinarily described as a weak paternalist because he is strongly opposed to paternalism towards competent adults, but in favour of it towards children and barbarians. However, there are two well-known passages in Mill which suggest strong paternalism. One of them presents Mill's ban on slavery contracts (On Liberty, p.125), and the other

his prohibition against long-term irrevocable contracts, such as marriage contracts which prevent the possibility of divorce (Principles of Political Economy, pp.953-4). It has been argued both that these passages are anomalies in Mill (Arneson, 1980), and, more plausibly, that they are reconcilable with Mill's overall position (Gray, 1983, pp.90-97), but either way Mill remains a staunch anti-paternalist towards competent adults.

Nearly all contemporary theorists on the topic are weak paternalists. Rawls, for example, says that the principles of paternalism are those which would be chosen behind the veil of ignorance in order to protect people against "the weakness and infirmities of their reason and will in society" (1971, p.249). "Paternalistic decisions", he says, "are to be guided by the individual's own settled preferences and interests in so far as they are not irrational, or failing a knowledge of these, by the theory of primary goods" (Ibid.). Rawls also emphasizes the subsequent free approval, or consent, of the recipient. We must be able to argue, he says, that when the recipient develops or recovers his rational powers, he "will accept our decision on his behalf and agree with us that we did the best thing for him" (Ibid). Other theorists reiterate similar themes. Dworkin emphasizes that consent seems the only acceptable way of trying to delimit an area of justified paternalism (1971, p.119). Feinberg (1971) argues that legal paternalism is never justified when conduct is voluntary or substantially voluntary, except when temporary intervention is necessary to establish whether it is voluntary or not. Similarly, Hodson (1977) argues that paternalistic interventions are justified

if and only if there is good evidence that the recipient's decisions are encumbered in some way, and good evidence that his decisions would be supportive of the intervention if they were not encumbered. Beauchamp and Childress (1979) and Childress (1982) argue in the context of medical paternalism that a state of at least partial involuntariness is necessary to justify paternalism, on the grounds that justified paternalistic interference is autonomy-respecting. The emphasis most theorists place on consent, voluntary choice, unencumbered wills, and so forth, indicates that a very high value is placed on personal autonomy.

The commonly-held belief that paternalistic interventions ought to be guided in some way by the principle of respect for personal autonomy has led most theorists to conclude that paternalistic interference with the fully informed, voluntary wishes of persons is never justified (qua paternalism), even if their choices are clearly self-harming and the subject would be grateful later for the intervention. This conclusion has sometimes led to the implausible view that children and others with less experience are incapable of acting in a fully voluntary way. It has led also to a questionable dichotomy in the usual justification of paternalism, stressing the priority of autonomy over well-being for adults and vice versa for children. Thus we need a developed theory of autonomy to show what constitutes its exercise, development, and infringement. Only then can we properly decide which paternalistic policies are autonomy-respecting and which are not. The main purpose of this study is to provide such a theory and to see what implications follow from it for autonomy-respecting paternalism.

Not everyone accepts the view that the principle of respect for autonomy has an important role to play in guiding paternalistic intervention. Husak (1981), for example, argues that we should assess each case of paternalism on its individual merits from a consequentialist perspective. He suggests that the acceptability of both paternalism towards oneself and paternalism towards others, when future consent has been reasonably assumed but not actually forthcoming, suggests that concern for the non-violation of the autonomy of the recipient need not feature in the justification of paternalism. However, cases of paternalism towards oneself are not the paradigm cases of paternalism with which we are concerned; and, as will be made clear, the acceptability of paternalism towards others when future consent is reasonably assumed but not actually forthcoming, does not tell against a theory of the justification of paternalism based on respect for the recipient's autonomy.

The starting point for the thesis is, therefore, the view that paternalism is morally permissible qua paternalism if it is both autonomy-respecting and beneficial to the recipient. Permissible paternalism qua paternalism may fail to be permissible from an all-things-considered point of view either because it is ruled out on non-paternalistic grounds, or because it conflicts with a paternalistic duty. Because the thesis is concerned only with delineating an area of permissible (autonomy-respecting) paternalism, it does not argue for any particular full theory of the good, from which paternalistic duties may be derived. It is argued in Chapter Three, however, that any full theory of the good from which paternalistic duties are derived, must be compatible with the theory

of well-being presented in that chapter.

B. Definition of Paternalism

Some people think of paternalism mainly as a model of government. As such it suggests a beneficent father figure guiding his children to maturity. But we are not concerned in this thesis with the model of the state as paternalist. Numerous other models of government are possible: the state as contractor making agreements with individuals or groups, the state as master in a master-slave relationship, the state as doctor in a doctor-patient relationship (the iatristic model), the state as people in conversation with one another, the state as representative of the people, as dominating friend, or even unpleasant mother-in-law (Fotion, 1979). Many of these are preferable to the paternalistic model of government. But in any case our concern is with paternalism independent of the model of government one adopts.

The absense of firm boundaries around the concept of paternalism is reflected in the wide variety of definitions adopted in the literature. Most theorists, however, include some reference to coercive interference in their definition. Dworkin, for example, defines paternalism as

interference with a person's liberty of action justified by reasons referring exclusively to the welfare, good, happiness, needs, interests or values of the person being coerced.

(1971, p.119)

This definition, which concerns only interference with liberty of action, leads to the consequence that if A interferes with the liberty of action of B, and A justifies the interference by reasons referring exclusively to B's values, then A interferes paternalistically with B whether or not A believes the interference will protect or promote B's well-being.

Other definitions avoid this counter-intuitive consequence. Arneson, in defending Mill's absolute ban on paternalism, reformulates Mill's principle about "the only purpose for which power can be rightly exercised" as:

Paternalistic policies are restrictions on a person's liberty which are justified exclusively by consideration for that person's own good or welfare, and which are carried out either against his present will (when his present will is not explicitly overridden by his own prior commitment) or against his prior commitment (when his present will is explicitly overridden by his own prior commitment), and paternalist policies so defined are always wrong. (1980, p.487)

This definition is tidier than Dworkin's, but it is also very restrictive. It excludes coercion from the class of paternalistic acts if the recipient's present will is explicitly overridden by his own prior commitment.

There are other less restrictive definitions. Carter defines a paternalistic act as one in which

the protection or promotion of a subject's welfare is the primary reason for attempted or successful coercive interference with an action or state of that person. (1977, p.133)

This definition includes certain acts of coercion excluded above, namely those in which the recipient's present will is explicitly overridden by his own prior commitment. An example is Odysseus's men refusing to untie him from the mast when they came within

hearing distance of the sirens. In addition, Carter's definition includes interference with a beneficiary's states, both mental and physical. It could be objected, however, that paternalism does not always constitute interference. Paternalistic lies, for example, or breaches of confidentiality, may not, strictly speaking, interfere with the recipient, and that is also true of paternalistic acts such as giving food vouchers instead of money to social security recipients when money would be preferred. Carter explicitly ignores such cases on the grounds that they are not the central cases of paternalism (1977, p.133). Dworkin explicitly ignores them too, although he suggests they may be defended on paternalistic grounds (1971, p.110).

Other definitions of paternalism easily accommodate paternalistic lies and breaches of confidentiality. Gert and Culver state that

- A is acting paternalistically toward S if and only if A's behaviour (correctly) indicates that A believes that
- (1) his action is for S's good;
 - (2) he is qualified to act on S's behalf;
 - (3) his action involves violating a moral rule (or will require him to do so) with regard to S;
 - (4) S's good justifies him in acting on S's behalf independently of S's past, present, or immediately forthcoming (free, informed) consent; and
 - (5) S believes (perhaps falsely) that he (S) generally knows what is for his own good. (1979, p.199)

Fotion argues that Gert and Culver's definition is an improvement on those we have mentioned, which define paternalism in terms of coercion or interference, but he claims it is still too restrictive on the grounds that not all paternalistic acts violate (even in any *prima facie* way) moral rules (1979, p.195). Providing services to one's children, he points out, can be paternalistic but need not

involve any violation of a moral rule. According to Fotion, A acts paternalistically towards B only if A acts with the intention of benefitting B, and A's action is one which B could, in principle, learn to perform for himself. On this account, regulating drug, chemical, and food manufacturers is not paternalistic because these practices require greater expertise, resources and time than the beneficiaries of these practices are capable of providing for themselves. On the other hand, providing food stamps for individuals or groups is paternalistic, since it is at least possible for the beneficiaries eventually to provide food for themselves.

Fotion's account has at least two drawbacks. First, it excludes, for no apparent reason, cases such as lying to someone on his deathbed from the class of paternalistic acts, because there is no action involved in such cases which the beneficiary could learn to perform for himself. Second, it fails to delineate an area of moral interest, since it includes such everyday events as lending a book to a friend (for that friend's benefit) among the class of paternalistic acts. Such acts violate none of the recipient's moral rights. However, the question of justification arises only where there is at least a prima facie violation of a right. For that reason it is important to restrict the definition of paternalism to apply to cases where there is a violation or at least a prima facie violation of a right.

Unlike Gert and Culver, however, we are primarily concerned in this thesis not with violations of any moral rule or infringement of a right by A towards B for B's benefit, but with violations for his own good of B's autonomy which is protected by various

prima facie rights to non-interference (as outlined in Chapter Two). It is possible that the theory put forward in this thesis could be extended to cover cases involving other rights, when they are violated for the right holder's benefit - two such cases are discussed in Chapter Four. But there are two important reasons (other than the fact that most paternalistic violations of rights are, or can be reduced to, violations of various rights to non-interference) for restricting our study to the justification of paternalism only where this involves a violation of autonomy-rights. The reasons are these. First, cases of the paternalistic violation of rights other than the right to non-interference, e.g. the right not to be deceived, cannot be identified until the prior question of which of these rights belongs to whom, has been settled. For this we need a complete theory of rights, which we do not have. Second, there may be some rights which are inalienable, unlike the right to non-interference which, it is argued in Chapter Two, is alienable by the right-holder. Even in giving up one's freedom, one exercises freedom in giving it up; so there is a sense in which one does not lose the benefit of non-interference even in forfeiting it. But that is not true of reception rights, e.g. the right to receive adequate food, clothing, shelter, and so forth. There is no sense in which one retains, in forfeiting them, the goods secured by reception rights.

For the above reasons, the following definition of paternalism is adopted in this study:

- A's act, P, towards some recipient, B, is paternalistic if and only if both
- (a) A's primary reason for acting is to protect or promote B's well-being, and

- (b) A prima facie violates B's (yet to be elucidated) right to non-interference.

C. The Importance of Autonomy

It is not part of the main purpose of this study to argue for the view that paternalism ought to be autonomy-respecting, but rather to show what autonomy-respecting paternalism entails, i.e. to show how paternalistic interference and respect for autonomy can be reconciled. Still, there are some general points to be made about the importance of autonomy, largely for the purpose of justifying this starting point.

It is commonly believed that there is considerable instrumental value in the possession and exercise of personal autonomy. Greater satisfaction and sense of achievement, for example, are often thought to be among the benefits that autonomous persons can hope to enjoy. This point is implicit in Rawls' discussion of the Aristotelian principle, which on his account states that other things being equal, human beings enjoy the exercise of their capacities, and this enjoyment increases the more the capacity is realized, and the greater its complexity. Rawls says:

Complex activities are more enjoyable because they satisfy the desire for variety and novelty of experience, and leave room for feats of ingenuity and invention. They also provoke the pleasures of anticipation and surprise, and often the overall form of the activity, its structural development, is fascinating and beautiful. Moreover, simpler activities exclude the possibility of individual style and personal

expression which complex activities permit or even require.
(1971, p.427)

If this is true, then since the successful performance of complex activities requires more knowledge and skill than the successful performance of simpler activities, and since knowledge and skill are required for autonomy (as shown in Chapter One), one might expect that autonomous persons would experience greater satisfaction than non-autonomous persons through their performance of complex activities.

Engaging in complex activities is not, however, the only source of satisfaction usually thought to be more available to autonomous persons. There is also the satisfaction of increased self-esteem which may accompany the development and exercise of autonomy. The same sense of dignity and self-worth which the self-directed person feels is not available in full measure to non-autonomous persons.

In addition, the development and exercise of autonomy is normally thought to be instrumental in producing benefits for society as a whole. "Mankind are greater gainers," says Mill, "by suffering each other to live as seems good to themselves than by compelling each to live as seems good to the rest" (On Liberty, p.17).

Besides the benefits to society of cultural diversity, and "experiments in living", if finding and implementing solutions to the many problems facing human beings requires expertise and initiative, then it might be thought that autonomous persons are better equipped to find these solutions than non-autonomous persons, thereby benefitting not only themselves but society as a whole.

The claim that the exercise of autonomy is instrumentally valuable, that it leads to other benefits, is, of course, defeasible. One could, perhaps, regard Huxley's *Brave New World* as a superior alternative to autonomous living, on the Skinnerian grounds that our notions of human dignity and moral agency are better regarded as sentimental anachronisms. Or one might simply have far less confidence than Mill that individuals know their own interests best. Hart, for example, says

Mill carried his protests against paternalism to lengths that may now appear to us fantastic. He cites the example of restrictions of the sale of drugs, and criticizes them as interferences with the liberty of the would-be purchaser rather than with that of the seller. No doubt if we no longer sympathize with this criticism this is due, in part, to a general decline in the belief that individuals know their own interests best, and to an increased awareness of a great range of factors which diminish the significance to be attached to an apparently free choice or to consent. (1963, pp.32-3)

Nevertheless, our primary concern is not with the instrumental value of autonomy but with its intrinsic value. For only if the possession and exercise of autonomy has intrinsic value of equal weight, as it were, to increases in well-being, can it be successfully conjoined with the latter in defining the limits of justified paternalistic interference. If autonomy is intrinsically valuable to us, this means that we value it for its own sake, independently of other goods such as happiness or accomplishment, whether or not we regard intrinsic value itself as mind-dependent, with Mill and Sidgwick, or with Moore, as independent of consciousness.

Feinberg's remark that "deriving and explaining the distinctive value of freedom is one of the great unfinished tasks of philosophy"

(1979, p.100), suggests the difficulty one feels in trying to show that autonomy has intrinsic value. There are no arguments to prove, of course, that anyone who does not value autonomy for its own sake is making a logical or a factual error. But there are many variations on two main lines of argument which suggest that autonomy is in fact very highly valued for its own sake.

The first line of argument emphasizes an intimate connection between autonomy and human dignity or personhood. Autonomy, according to Kant, is a matter of acting on self-legislated universalized laws. His conception of autonomy is very different from that presented in Chapter One, but his theme of an intimate connection between autonomy and human dignity or personhood is reiterated in much contemporary philosophical (and non-philosophical) literature.

Downie and Telfer, for example, suggest that personhood implies the capacity to be self-determining and rule following (1969, p.35), i.e. that personhood implies, to some degree, personal autonomy. Similarly, Benn argues that we are committed to respect the person who is "his own cause, his own handiwork" in so far as we care about the ontological status of persons, on the grounds that autonomy is constitutive of personhood (1976, p.129). And Griffin states that

choosing one's own course through life, making something of value out of it according to one's own lights, is what makes one's existence human. If you value your humanity, you will value what makes life human, over and above what makes it happy.
(1981, p.59)

The fact that a strong connection between autonomy and human dignity

is widely assumed to hold, in itself suggests that autonomy is one of our most fundamental values, perhaps the most fundamental value.

The second line of argument normally takes the form of a "thought experiment" in which we are asked to consider how we might react in certain hypothetical cases, with the aim of isolating our fundamental values. Exploring our reactions to various kinds of experience-machine is a particularly useful thought experiment in the context of focussing on the intrinsic value of autonomy. According to this experiment, we are to imagine that technology is sufficiently advanced so that we have a machine at our disposal which would provide us, for life, with the sensation of having any kind of experience we might choose. Presumably, most of us would not plug in, and we are asked to consider why not. Nozick (1974, pp.421 and in greater depth Glover (1984, pp. 92ff) discuss the values behind the refusal to plug in. It seems that what we value most is living our own life. This involves both contact with reality and controlling our life, in spite of our obvious willingness to make temporary, intermittent, or even permanent use of various kinds of machines in satisfying many of our desires.

Glover points out that reality and control are important to us because we value three things most: genuine contact with other people, the opportunity for expansion of consciousness, and self-development and self-expression (1984, Part II). Reality is needed for the first and second of these, and control for the third. There may be disagreement about what it is to be in control, or about what kinds of influence count as manipulation by others,

but we generally agree that the idea of control by others including machines, is morally repugnant. The problem of paternalism is the problem of how to reconcile our allegiance to this fundamental value of self-government with our desire (and sometimes our duty) to help protect others from their own self-harming conduct.

D. Justification

The view that paternalism is justified if and only if it enhances the recipient's well-being and does not infringe his autonomy is presented in this study primarily in terms of the objective states of affairs which a paternalist committed to the principle of respect for autonomy would, if he knew of them, approve. The focus is on autonomy, rather than well-being, although the latter is the subject of Chapter Three. It is important to focus on an "objective-conditions" sense of justification in this context, as opposed to the broader good-reasons-for-acting account, because it is very easy for a paternalist to have good reasons for acting which give the wrong result.

For example, suppose that on the day of Arthur Koestler's death his housekeeper had arrived for work unusually early and found Koestler and his wife still alive, after taking the overdose of tablets which actually killed them. Suppose the housekeeper, knowing nothing of Koestler's well-planned decision to end his life, had telephoned for an ambulance which rushed Koestler

to a hospital where, against his instructions, the doctors pumped his stomach and revived him. It is very doubtful that this paternalistic intervention on the part of the housekeeper and the doctors would have provided any autonomy-respecting benefit for Koestler. We have good evidence for believing his suicide was rational and well-planned. Because the interference would not have resulted in autonomy-respecting benefit for Koestler, it would not have been objectively justified, in spite of the fact that both the housekeeper and the doctors would have been justified in the sense of having good reasons for acting as they did in the circumstances.

It is argued in Chapter Two that an agent can alienate a prior prima facie right to non-interference by subsequently approving (in a technical sense specified in that chapter) of the prior prima facie violation of that right, thus rendering it a non-actual violation. Several reasons are given there why subsequent approval is the appropriate vehicle for this. However, it is sometimes suggested that justification for acting paternalistically cannot depend on anything as unpredictable as subsequent approval, (by for example Ten, 1980, p.112; and Lively, 1983, p.153). The worry behind this claim is readily understood. One does not want to be accused of acting unjustly when consequences are unavoidably unpredictable. But this worry is unfounded, for justification in the objective sense, as we have called it, is largely a separate matter from the blameworthiness or praiseworthiness of one's reasons for acting. At the same time, it does not follow from the uncertainty of future occurrences that they should not feature in moral thinking, either consequentialist or non-consequentialist. Thus, in this

study it will first be decided what states of affairs are consistent with autonomy-respecting benefit (the objective conditions of justification), and then our reasons for acting can to some extent be tailored to fit accordingly.

There are three "levels", as it were, of justifying paternalism. They can be separated by asking three questions:

- (1) What are the objective conditions of justified paternalism?
- (2) What moral rules should we adopt regarding paternalistic intervention to ensure that those objective conditions will be met?
- (3) What institutional rules (e.g. in schools, hospitals, and courts of law) regarding paternalistic intervention should there be?

These three questions form a hierarchy which cannot be answered all at once. The answer to (3) depends on the answer to (2), which in turn depends on the answer to (1). Most writings on paternalism are concerned with questions (2) and (3). The contention in this thesis is that (2) and (3) cannot be fruitfully pursued without first answering (1). At the same time (1) and (2), and (2) and (3) are very closely interconnected. But, providing a comprehensive answer to (1), and seeing at least in part what follows from it, is the aim of this thesis.

The rules at levels (2) and (3) have point in a way that is analogous (but only roughly) to the way in which a rule utilitarian's rules have point. Following the moral and institutional rules is supposed to lead, in general, to the states of affairs which constitute the objective conditions of justified paternalism. The analogy is only very rough, however, for there is not meant to be the same kind of solid divide between

levels (1) and (2) as there is, for example, between Hare's (1981) two levels of moral thinking. Although the answer to question (2) depends on the answer to (1), the influence also goes the other way. Level (1) is concerned with the conditions of autonomy-respecting paternalism, but it is at level (2) thinking in the first place where it is decided that paternalism should be autonomy-respecting.

Another way of putting the point that the focus of the thesis is on the objective level of justification, is by drawing attention to the distinction between what an autonomy-respecting paternalist has most reason to do, and what, given his beliefs, it would be rational for him to do. Parfit makes this distinction in a different context using the following example: "If my wine has been poisoned, drinking this wine is not what I have most reason to do. But, if I have no reason to believe that it has been poisoned, I would not be acting irrationally if I drink this wine" (1984, p.153). Arthur Koestler's housekeeper would not have acted irrationally in calling an ambulance, nor would the doctors have acted irrationally in pumping his stomach, but given the particular circumstances of the case, interventions like these are not what autonomy-respecting paternalists have most reason to do. Our focus in this thesis is not on what it would be subjectively rational to do, but on what an autonomy-respecting paternalist has most reason to do, regardless of his beliefs.

There is of course ultimately no point in isolating the objective conditions of autonomy-respecting paternalism unless our reasons for acting in particular circumstances change to fit

those conditions, but isolating the objective conditions is the prior task. Afterwards, we shall see that someone's being informed, competent, and acting voluntarily is, surprisingly, not always a good reason for refraining from paternalistic interference with him - and this is so without embracing strong paternalism.

E. Overview

A theory of autonomy is presented in Chapter One. There the importance of rational reflection to personal autonomy is discussed in some detail. It is argued that there are two aspects to the exercise of autonomy besides liberty, namely self-construction and self-control. Self-construction is the activity of controlling, as far as that is possible, one's most fundamental beliefs and desires through casually efficacious reflective input into their formation. Knowledge, including self-knowledge in the sense of strong evaluation, and moral sensitivity, are central to self-construction.

Self-control is discussed as the activity of bringing unruly beliefs and desires into line with the central core. Unruly beliefs and desires are those which are either logically or empirically incompatible with fundamental desires and beliefs. Some failures of self-government such as self-deception, neurosis, and weakness of will, are regarded as failures of self-control. Others, such as anomie and ambivalence are taken as failures of self-construction. At the end of the chapter, three objections

to the account of autonomy we adopt are considered.

The right to non-interference, the subject of Chapter Two, protects liberty, self-construction, and self-control. The right to non-interference can be divided into at least three rights of greater specificity, each of which can be divided in turn into a large number of highly specific rights. The right to non-interference₁ is the right to non-interference with regard to actions. The right to freedom of speech is a specific instance of it. The right to non-interference₂ is the right to non-interference with with mental and physical states, and non-interference₃ is non-interference with an agent's opportunities. The source and extent of these rights, as prima facie rights, and their alienation by the rights' holder is the subject matter of the chapter.

The source of the prima facie right to non-interference is grounded in the intrinsic value of autonomy, and its possession is extended to all beings with desires. It is argued that a prima facie violation of a right to non-interference is not an actual violation if and only if the right holder (freely or autonomously) subsequently approves_e (technical term) of the prima facie violation. If the prima facie violation is a deprivation of liberty, the recipient's subsequent free approval_e alienates this right. But if the prima facie violation inhibits self-construction or self-control, then subsequent free approval_e is not sufficient, and minimally autonomous approval_e is required.

'Approval_e' refers to approval in the (extended) sense of having an overriding desire that some past act of interference, i, should have occurred. The overridingness of a desire that i should have occurred, is a matter of both the strength and persistence of the desire (which may or may not be conscious or

reflected upon), and its hierarchical position within the agent's whole subsequent network of fundamental desires. This mechanism for alienating the right to non-interference is defended on independent grounds, and on the grounds that it is consistent with the exercise of autonomy as self-construction.

Chapter Three is about well-being. Although the thesis is primarily about autonomy and paternalism, justified paternalism always enhances the recipient's well-being. Well-being warrants some consideration for this reason alone, but also for another reason. An individual's well-being is understood in this chapter as the satisfaction of his informed, fundamental desires. This has some important implications for autonomy-respecting paternalism. It is argued that unless paternalistic interference contravenes some deeply held religious or moral principle of the recipient's, large increases in fundamental desire satisfaction resulting from paternalistic interference will automatically tend to be coincident with the alienation of the recipient's right to non-interference. Thus they will be autonomy-respecting. From this it follows, that there can be no large paternalistic gains in well-being for a recipient at the price of some small cost to his autonomy. This view defeats the commonly held position that it is sometimes appropriate to sacrifice a little autonomy for a large gain in well-being in paternalistic intervention, particularly towards children.

In addition, a place is found in Chapter Three for objective-list accounts of human well-being (as one source of paternalistic duties) provided the items on the list satisfy informed fundamental desires. It is emphasized that a definition of well-being in

terms of the satisfaction of informed desires, is consistent with the view that some informed desires are more in keeping with human flourishing than others.

Chapter Four sets out several implications of the foregoing chapters for paternalistic intervention. General implications concerning incompetence, consent, and children are considered, and specific cases of inter-personal paternalism and legal paternalism are discussed.

CHAPTER I

AUTONOMY

There is no consistent usage of the term 'autonomy' either in ordinary speech or in philosophical discourse. Here the term will be used interchangeably with 'self-government'. It is defined in IA below, but first it is worth mentioning two familiar distinctions which help to bring the concept to be elucidated more sharply into focus.

In the tradition of Rousseau and Kant, someone is autonomous in so far as he acts on self-legislated principles. On a narrower view, usually associated with Mill (although Mill did not use the term 'autonomy' in his essay On Liberty), a person is self-governing in so far as he is not forcibly prevented from acting on his voluntary choices. On this view, an agent's voluntarily selling himself into slavery diminishes his freedom but not necessarily his autonomy.¹ Certain aspects of the distinction between these two ideas is the subject of Berlin's (1969) famous discussion of positive and negative liberty. Sometimes the idea of not being prevented from acting on voluntary choices is referred to as freedom (or autonomy) of action, and the Kantian ideal of acting on self-legislated principles is referred

to as freedom (or autonomy) of the will. That is the first distinction.

Freedom of action, particularly civil liberty, is sometimes discussed as if it were unrelated to freedom in the sense of formulating self-legislated principles. Scruton (1983), for example, argues that the latter is more important to the integrity of the person, and the state should be more concerned with its development than with the promotion of mere civil liberties. Others stress the priority of freedom of action over autonomy of the will, on the assumption that the latter somehow follows from the former. However, both autonomy of action and autonomy of the will are regarded here as important requirements for the fullest possible exercise of self-government. On its own, freedom of action seems pointless if persons are unable to formulate self-legislated principles on which to act. On the other hand, the formation of self-legislated principles also loses point if agents are not free to act on them. Both the civil theorist and the "Promethean", as Cooper (1983) calls the (neutral) advocate of positive freedom, are concerned with different sides of the same coin. Both are concerned with overcoming those factors which inhibit an agent's ability to formulate and put into practice aims and purposes of significance to him. Hence, both are included in our account of self-government. Both generate, as we shall see, important rights to non-interference.

Autonomy of will is best described as an ideal of character. It is elucidated in IA2 and IA3 below as a neutral (as far as

that is possible) ideal - neutral in so far as it recommends no particular content for the beliefs, desires, aims, actions, values, attitudes, dispositions, and so forth, of self-governing persons. (The place of non-neutral ideals of character is briefly discussed in IB2 and Chapter III.)

The second distinction is between autonomy and free will in the sense required for moral responsibility. Much of the traditional literature on freedom is about metaphysical issues, such as the conflict between free will and predestination, which are clearly irrelevant to the practical account of autonomy with which we are here concerned. But it is also about issues which closely affect self-government, e.g. compulsion, securing conformity between first and second order desires, and so forth. What is the relationship, one might ask, between autonomy and the freedom required for moral responsibility?

The point has frequently been made that freedom is a necessary but not a sufficient condition for autonomy (Peters 1973, Benn 1976, Young 1980a). Young, for example, says:

One must be free to be autonomous, but one can be free and still lack autonomy because it is, for instance, possible freely but mindlessly to mimic the tastes, opinions, ideals, goals, principles, values and preferences of others. So freedom is necessary for autonomy but not sufficient. At least this much more is needed: that a person's choices actually be expressive of his or her individual preferences, aspirations and so on. (1980, p.35)

In apparent contrast, Dworkin (1981) argues that freedom of action is not necessary for autonomy, on the grounds that the drug addict who does not mind being an addict is to some extent autonomous, even though he is not free to act differently. In

an earlier paper, Dworkin (1970) argues that freedom of action consists in acting for reasons one does not mind acting from, but he later argues that this second-order preference theory of freedom seems to apply to a broader notion than acting freely. "For it is clear," he says, rightly, "that we resent other ways of being influenced besides coercion" (1981, p.210). Because we mind being motivated by false beliefs, by manipulation, by being kept uninformed of relevant information, the earlier analysis is incorrect. (Deceiving a dying patient about his condition, he says, is an interference with the patient's autonomy, but not his freedom.) Dworkin is right, for the reasons he gives, that a second order preference theory applies to a broader notion than acting freely. However, his example of the compelled but willing addict does not show that freedom in the sense required for moral responsibility (whatever that amounts to) is unnecessary for autonomy. The addict in Dworkin's example could not give up his habit even if he wanted to, but he is free to be either a willing or an unwilling addict. Freedom is a necessary condition for autonomy - freedom, that is, in the sense required for moral responsibility.

Like Dworkin, several writers have offered analyses of freedom which are divorced from the traditional debate about determinism and moral responsibility, and are more like partial analyses of the concept of autonomy presented later in this chapter. Frankfurt (1971) argues that a person exercises freedom of the will in securing conformity of his will to his second-order volitions, and suggests that this is all the

freedom it is possible to desire or conceive. This is important for distinguishing, as Frankfurt does, between a person and what he calls a 'wanton' (an agent lacking second-order volitions), but it leaves open the possibility that a compulsive killer acts freely as long as he is content to have the will he has. In a later paper (1975), Frankfurt alludes to the importance of maintaining a closer connection between the notions of freedom and moral responsibility than the earlier paper suggests; but in any case, having the will one wants to have is more like self-control than either the freedom required for moral responsibility or all the freedom it is possible to desire.

Watson's (1975) analysis is similar. Although he rejects Frankfurt's second-order volition theory, Watson argues that an agent is unfree to the extent that he is unable to get what he most wants or values, and this inability is due to his own motivational system. This, too, suggests the idea of self-control, and it also leaves open the possibility that a compulsive killer acts freely as long as he has the will he most wants to have.

Glover improves on these accounts by defining the ideal case of freedom more fully as one in which three conditions hold:

There must be nothing preventing or hampering the expression of my decisions in action. There must be harmony between my values and the desires which dominate my choices. And there must be harmony between both of those and my project of self-creation.
 ("Self-Creation", forthcoming.)

The latter two conditions point to a fuller notion of self-government than either of those discussed by Frankfurt or Watson, and the first also points to some aspects of the freedom that is required for moral responsibility. Both freedom in the sense required for moral responsibility, whatever that amounts to, and freedom in the sense required for the character-ideal of self-government, are aspects of 'all the freedom it is possible to desire or conceive'; but we are concerned here only with the latter of these, and the former is taken as a necessary condition of it.

A. Full Autonomy

1. Liberty

Autonomy of action, or liberty, in the sense of not being forcibly or coercively prevented from acting on one's voluntary choices, is, fairly obviously, an important requirement for the effective exercise of full self-government. But it is not the most important requirement, for someone might be a highly autonomous thinker while imprisoned, or alternatively, largely heteronomous while enjoying full liberty. Here we simply list it, as one of the major aspects of full autonomy, and set it aside until Chapter Two where there will be more to say about it in relation to the right to

non-interference.

It is sometimes thought that the problem of justifying paternalism involves only infringements of liberty of action. This assumption is implicit in the first definition of paternalism discussed in the Introduction (p.6). But respect for autonomy involves much more than non-interference with actions. Interference with an agent's beliefs, desires, states, opportunities, or powers, can equally infringe his autonomy rights. Liberty of action is perhaps the most frequently infringed autonomy right, but it cannot stand alone in forming the nucleus of an anti-paternalist stance based on the principle of respect for autonomy. An agent's beliefs, desires, intellectual development, and so forth, are as much a part of him as his actions, and like actions, they are the appropriate objects of a prima facie claim to non-interference from others. As we have said, the various rights to non-interference which protect full self-government, including liberty of action, are the subject of the next chapter. Suffice to note here that liberty of action generates important autonomy rights, but not the only autonomy rights. Other rights are generated by the other two components of autonomy: self-construction and self-control.

2. Self-Construction

In the following pages the terms 'governing self', 'governed self', 'fundamental self', and 'peripheral self' are

used. Before proceeding, their meanings should be clarified.

Where there is self-government, there must be some identifiable self which acts as governor, and other aspects of the self which are governed. Normally, the governing self consists of settled principles, convictions, aims, allegiances, and so forth; and the governed self consists of less settled and less central parts of the person, such as transitory desires, impulses, attractions, tentative beliefs, and so on. Feinberg refers to these two aspects of the self as the inner-core self and the wider self:

The attributes of the inner core self that rule the wider self include not only 'Reason', but materials for reason to work with, and the latter will be the convictions, ideals, and purposes that are most deeply entrenched in a hierarchical network of similar principles, those which, because of their logically central position in the network, are the last to be tampered with when changes of mind or heart must be made.

(1973, p.160)

This model provides us with a very rough account of how it is that the same self can be both governor and governed. It is important that the inner core self be composed of fairly settled principles, aims, beliefs, purposes, loyalties, etc., for if they were not fairly settled and constant, we could not recognize a continuing, cohesive self within a person. However, Feinberg's view that the inner-core self is made up of convictions, ideals and purposes which occupy a logically central position in the network seems unrealistic. The convictions, ideals and purposes of the inner-core self clearly occupy a central position, but they do not always occupy a logically central position. An aim to complete one's doctoral thesis, for example, within a

specified length of time, may be more central than, and incompatible with, a desire to spend six months touring India, but it is not necessarily logically more central. It may be simply more important.

Nevertheless, an autonomous agent's most important beliefs and desires do not exist as a set of isolated fragments, but as a (more or less) integrated whole. The cognitive, affective, and volitional elements of an autonomous agent's convictions, ideals, purposes etc., are closely interdependent, and fit together in some recognisably coherent way. In this loose sense, there is indeed a kind of logical hierarchy in the autonomous agent's whole network of beliefs and desires.

An agent's beliefs and desires which are most important to him, those which best reflect his deepest values, will be referred to as his fundamental beliefs and desires. Fundamental beliefs and desires normally include an agent's views and feelings about himself and his relationships with others, his beliefs and attitudes concerning God, human nature, the universe, morality, aesthetics, the "meaning" of life, the meaning of his life, and so forth, however well or poorly he is able to articulate those views. They are the beliefs and desires with which the agent most closely identifies. In contrast with fundamental beliefs and desires, peripheral beliefs and desires are logically and effectively less supportive of the agent's whole network of beliefs and desires. Usually an agent feels no great sense of loss or shift in his identity when changes are made in his peripheral beliefs and desires.

Unconscious beliefs and desires, however powerful, find no place in the distinction between fundamental and peripheral desires and beliefs. They are accounted for in terms of the distinction between the governing and the governed self. The governing self consists of those beliefs and desires which in fact best explain both an agent's actions and his other beliefs and desires. Thus the fundamental self and the governing self need not coincide. The governed self consists of the agent's beliefs and desires which, although they might explain some actions, are subordinate in explaining behaviour to other beliefs and desires, either consciously or unconsciously held.

It is important to emphasize that in the autonomous person the fundamental self is also the governing self. That is so because the autonomous person is self-controlled, i.e. both his peripheral and unconscious beliefs and desires are subordinate to his fundamental ones. Self-control is considered in IA3 below. It will be argued throughout IA2 that in the autonomous person the fundamental self is self-constructed through rational reflection.²

(a) Rational Reflection

i. Stability and Continuity. Let us consider the hypothetical situation of two persons, A and B, both of whom have the same conviction that, let us say, there are objective moral values independent of human desires. In both cases the conviction has played a central role in determining the admissibility of other elements into the subject's whole

network of beliefs and desires. In both cases the conviction is central and apparently settled. The difference between A and B is that A adopted his belief in the objectivity of moral values in an unreflective manner from parents who held the same deep, unreflective conviction as part of their religious outlook. B, on the other hand, adopted his belief after examining the arguments for and against both positions. B carefully considered the views of Mackie, McDowell, Blackburn, and others on the matter before deciding which view to adopt. Although both A and B hold settled convictions, we regard B as more self-governing than A. Why?

It is clear that we generally admire choices such as B's, where the options have been reflected upon, weighed and carefully considered, more than we admire choices based on either convention or impulse. There are good reasons why this should be so. Acting unreflectively is more likely to prove harmful to the agent or others, than acting after weighing relevant considerations. We often regret unreflective acts, sometimes because of failure to give due weight to the interests of others, and sometimes because unreflective acts turn out to be unrepresentative of our deepest preferences and beliefs. However, the fact that we generally trust reflective choices more than unreflective choices, does not in itself illuminate any connection between rational reflection and self-government.

It could be thought that we are inclined to regard B's belief, but not A's, as more his own, rather than simply his, because we can point to particular persons, namely A's parents,

who influenced A's views, about the objectivity of moral values. The view in question is not A's own, it might be thought, because it really belongs to his parents, or perhaps his grandparents. But suppose we could point similarly to some influential person in B's life, who strongly believed in objective moral values, but who urged B to examine the viewpoints of several different writers on this topic. The fact that we could point to particular influential persons in A's or B's life, makes little difference in itself. We are not immune, nor would it be desirable to be immune, to the influence of parents, teachers, friends, writers, and so forth. Of course sometimes the influence of one person on another is "undue" influence, in which case we are less inclined to count the beliefs and desires acquired as the result of such influence as the subject's own, than we are with beliefs and desires acquired in some less objectionable way. But ex hypothesi we are not dealing with cases of undue influence. So the fact that there are influential others provides no explanation of why B's belief in objective values should be more his own than A's.

A more promising line of explanation for regarding B's conviction, but not A's, as his own, is that B struggled much more than A to arrive at his belief. B worked harder. Perhaps the initial plausibility of this idea rests on the view that if someone has put greater mental energy into acquiring a belief, if he has "invested" considerable intellectual and emotional effort, then he has a greater claim to ownership. This view has some intuitive appeal. However, it seems to rely on some

kind of "labour theory" of ownership which in no way captures many of our central intuitions about ownership in general.

B can have a house of his own without investing his labour in it, so why not an idea? We do not deny his claim to ownership simply because the house was acquired from parents who owned it first.

On the other hand, it is possible that a labour theory of ownership might fail with regard to material goods, while retaining plausibility with regard to beliefs and desires. But the crude version emphasizing only quantity, and not quality of reflective investment must be modified for the following reason. Someone could reflect on a belief at very great length, but in the manner of a brainwashed or indoctrinated person without thereby making the belief his own. Therefore, his reflection must be rational in the sense that he adopts (or retains) beliefs, convictions, allegiances, etc. for reasons. The reasons must be causally efficacious; that is to say, they must not be rationalizations. Provided an agent's reflection is rational in this sense, a "labour theory" of ownership can be plausibly applied. But still the connections between having ideas of one's own, as opposed to simply having them, and self-government, remain to be spelled out.

First let us consider why rational reflection is important to self-government, with the emphasis on the first component of that hyphenated term. This can be brought out by considering again the governing self. The governing self normally consists of settled beliefs, desires, evaluations, and so forth. Even though A's belief in objective values may seem

to be settled, what confidence can we have that it is actually settled unless it continues to be held only after alternative beliefs, and the grounds for holding them, have been considered. Perhaps A would give up his belief in objective values if he seriously considered Mackie's "queer entity" argument, or the argument from relativity. Or perhaps his belief would be weakened if it were pointed out to A that he simply adopted this view as a child on his parents' authority, in spite of the fact that his parents could not have defended it very well either. We cannot be sure that A's belief is a stable part of his governing self unless the belief is held by him after weighing up various counter arguments. Since B has rationally reflected on the pros and cons of his position, we have reason to be more confident that his conclusion, whatever it is, will stick. Thus, rational reflection both helps ensure stability and provides evidence, as it were, that a belief is a more stable part of the governing self than an unreflective belief.

That might be true of beliefs, it could be thought, but does it also apply to desires? A person can reflect on his beliefs before adopting them as his own, but he cannot as easily reflect on what it makes sense to desire before desiring it. If desires are reflected upon, are they more one's own and less vulnerable to change than if they have not been reflected upon? It may seem that unlike beliefs, desires are less affected by reflection. Beliefs, after all, are grounded in evidence in a way that desires are not. However, we know that sometimes desires change upon reflection on their causes, their

consequences, or the worth of their objects. Also, although desires change as beliefs and circumstances change, desires about the kind of person one wants to be, and the kind of life one wants to live, i.e. desires that most affect one's life, are generally more stable when they persist in spite of reflection on alternative ways of life, various options that might be open, knowledge of how these desires were acquired, and so forth. Again, rational reflection helps ensure stability and also provides evidence that desires of this kind are more stable within the governing self than unreflective desires.

It is important that the beliefs and desires which constitute the governing self be stable to some degree (quite apart from the psychological value of stability), for as we have said, without stability we could not recognize a particular cohesive self. But the importance of stability should not be overemphasized. The governing self should be conceived as a dynamic, ever-changing central core, and not as a static collection of beliefs and desires. This brings us to the point that rational reflection is related not only to the stability of the governing self, but also to its continuity. If there is to be a single governing self, then it is important that subsequent stages of its development be recognizably continuous with former stages. Rational reflection is the vehicle of continuity. It relates earlier beliefs and desires to subsequent ones.

But rational reflection contributes much more than stability and continuity to the governing self. When conjoined

with deep evaluation it also gives "substance" or depth to the governing self. More is said about that in IA2(b) below. Now we turn to the role of rational reflection in self-government, with the emphasis on the second component of that term.

ii. Control. The most important connection between rational reflection and self-government is not so much its contribution to the stability and continuity of the governing self, but the role it plays in controlling the governing self. The difference between A and B in the example of the previous section, is that B has taken an active part in forming his governing self, whereas A has been a passive recipient. This is important because only through the activity of rational reflection can an individual exercise some degree of control over the kind of person he is and will become. If an agent has some control, through reflective input, over the fundamental beliefs and desires which constitute his governing self, then he is more self-governing than he would be if his influence extended only over the peripheral parts of himself.

If, for example, someone is firmly indoctrinated from an early age into certain religious beliefs, and those beliefs become an unquestioned part of his adult governing self, it may be possible for him to exert enormous self-control in ridding himself of any tendency to embrace beliefs or desires in opposition to his central core, but in such a case his control amounts to little more than protection of a passively received governing self. If this agent were to reflect at some length on the arguments for and against the beliefs he holds, seriously

questioning them in the light of arguments advanced by others, and taking into account the fact that his beliefs were originally accepted not for reasons which he himself adopted as relevant, but because he was indoctrinated at an early age, he might, after weighing the arguments, give up his religious beliefs and adopt others in their place. Or he might decide that, in spite of past indoctrination, his beliefs are soundly based. If he makes the latter decision, it may be difficult to decide whether his apparently sincere (causally efficacious) rational reflection is actually rationalization, but the more reflective he is, the more he will take that possibility into account. By reflecting on various beliefs and desires before incorporating them into his fundamental core, the agent comes to accept some and reject others, thereby actively constructing his fundamental self.

Of course he does not do this in a vacuum. Reflection could never get off the ground if all beliefs and desires were questioned at once. Questioning can occur only by using some considerations as settled criteria against which other considerations can be measured. But this does not blur the distinction between the agent who has a hand in determining his most fundamental beliefs and desires, and the agent who simply adopts them passively from some external source. The present point is, simply, that the agent whose fundamental beliefs and desires are the product of, or survive the process of, rational reflection, is exercising more control over his governing self than the agent who is passively acquiescent about them.

It could be objected that the control the reflective agent exercises over his governing self is something of an illusion. Reflective input, even causally efficacious reflective input, should not be equated with control, so the objection might go, on the grounds that our reflections are themselves controlled by a variety of physical and psychological causes. But how plausible is this objection? As Nozick points out, no one has ever announced that because determinism is true, thermostats do not control temperature (1981, p.315). Even if the notion of control is transitive, which it seems to be, that fact does not detract from the control that may be exerted by an agent over himself. If A controls B, and B controls C, then A also controls C; but the fact that A controls C does not negate the fact that B controls C as well. That is to say, it does not follow from the fact that an agent's reflective reasonings are caused by forces outside his control, that an agent can exert no control over his central beliefs and desires. Perhaps we are used to thinking of self-government in terms of bringing ourselves to act in accordance with certain fundamental beliefs and desires to which we are already committed (usually referred to as self-control), but the control exercised by the self-constructed agent is control over the commitment to these fundamental elements themselves.

It could also be objected that the control the reflective agent exercises over his governing self is something of an illusion, on the grounds that if A controls B there must be some particular state of mind, or a state of mind from within a

particular range of states of mind induced by A in B, even if A and B are one and the same person (Dennett, 1984). According to this objection, if an agent rationally reflects on arguments for and against various religious beliefs before accepting or rejecting them, he is not really controlling his beliefs, for there is no particular belief state, or no belief state from within a particular range of belief states that the agent induces in himself. At least three rebuttals can be made to this objection. First, it could be pointed out that the objector's concept of control is unduly restrictive. We often speak of factors such as rainfall, temperature and nutrients in the soil controlling the vegetation of a region; or again, we often speak of thermostats controlling temperature, without any reference to particular states of mind in either the controller or the controllee. Used in this broad sense, the notion of control functions something like the notion of influence. Second, if the agent sincerely weighs up reasons for and against retaining certain religious beliefs, then he is committed to inducing in himself whatever particular belief state is appropriate to the truth of the matter. According to this rebuttal, no more particularity concerning state of mind is needed for the objector's restricted sense of control to apply. Third, if it is objected that the use of the term 'control' is inappropriate in characterizing the distinction between the person whose fundamental beliefs and desires are passively received, and the person whose fundamental beliefs and desires are actively constructed, some other term might easily be used to mark this distinction; but in any case the distinction remains intact.

Finally it could be objected that control of the governing self is an illusion on the grounds that a belief in the value of rational reflection, or a desire to reflect first before adopting various beliefs and desires as one's own, is itself a fundamental commitment which is normally passively acquired in the course of one's upbringing. Since it is passively acquired, it is a foreign element of a self-constructed governing self. But that does not follow. A commitment to reflection may be passively acquired, but reflective people are generally well aware of both the intrinsic and extrinsic value of reflective activity. Other kinds of passively maintained beliefs and desires are not like that; their possessor is usually unaware of arguments for and against commitment to them and to various alternative beliefs and desires. But even if our commitment to rational reflection were itself wholly unreflective, that would not prevent us from adopting rational reflection as the only available vehicle of control of the governing self by the self.

At this point someone might object that although it may be possible to be "self-constructed" through rational reflection, one need not take the reflective route in becoming self-constructed. One could, so this objection might go, simply choose certain elements to be included in one's governing self, based on their intuitive appeal, and set out to acquire them. But this is a naive objection, for we are not so malleable as to be able to choose our most fundamental beliefs and desires in

this simplistic way. It is possible for an agent to decide in advance to acquire a taste, say for Bach, or perhaps a belief in God, with the idea that once acquired this taste or belief will form part of his governing self. (This is the project of self-creation discussed in IA3(c) below.) However, there must be certain prior beliefs and desires in the governing self to begin with, from which follows a commitment to acquire a taste for Bach or a belief in God. It is the way in which those beliefs and desires have been rationally reflected upon which is the measure of the extent to which the governing self is self-constructed.

Rational reflection we have said, must not be rationalization. It must be causally efficacious in maintaining one's fundamental beliefs and desires. There are two contributing factors which enhance the quality of reflective ability: knowledge of the kinds of things there are to reflect about, and acquisition of the skills one needs to reflect with. These we may call the components of rational reflection.

iii. Components. If knowledge could be neatly divided up into a manageable number of different kinds or forms of knowledge, we would have the welcome convenience of being able to specify in advance the areas in which one must be immersed, at least to some extent, in order to reflect well. It has been argued (by Hirst, 1972, pp.15ff) that there are only a small number of basic forms of knowledge - mathematics, physical science, human science, history, religion, literature and the fine arts, philosophy, and ethics - on the grounds that each of

these has its own central concepts, logical structure, distinctive expressions, and distinctive techniques for exploring experience and testing those expressions. Fields of study such as geography or engineering supposedly draw on several of the basic forms. According to Hirst these derivative disciplines do not manifest any one logically distinct form of expression. Hirst (1974) later drops History from the list of basic forms of knowledge, and in place of human sciences he adds knowledge of one's own and other minds. But the exact content of Hirst's list is not important for there are several reasons why we cannot expect any such list to help us in the context of trying to determine whether there is a non-random approach an agent might take towards increasing his overall capacity to reflect by increasing his knowledge.

First, the arguments for this kind of reductive view are unconvincing. It is not clear that each area of discourse has distinctive expressions that are testable against experience in accordance with particular criteria that are peculiar to the form. What are the criteria peculiar to religion, or literature and the fine arts? Hirst admits that some forms of knowledge, particularly moral knowledge and the arts, are such that what the tests are and the ways in which they are applied are only partially statable (1972, p.16). He adds that some critics would dispute the status of these forms of knowledge as basic forms for this very reason. However, as we know, even the status of the physical sciences can be disputed on similar grounds. Kuhn (1970), Rorty (1980), and others present highly controversial

views suggesting that the authority of the tests against experience for the expressions of science can be called into question. If this authority depends in large measure on consensus, as Rorty emphasizes, Hirst's views or others like them face the same criticism that is made of Wittgenstein's forms-of-life doctrine, namely, the threat of an excessive proliferation of forms of knowledge wherever there is consensus about what counts as a test for distinctive expressions.

Second, even if we had a non-controversial list of basic forms of knowledge, it would be of little practical help in the pursuit of increased autonomy because there would still remain the problem of whether an agent can reflect better if he achieves breadth in several areas rather than depth in one or two areas, given that either breadth or depth must be sacrificed one to the other. Also, it is not the case that immersion in each of a number of distinctive forms of knowledge would enhance an agent's ability to reflect on issues that most affect his particular life. Knowledge of mathematics, for example, may be less relevant for general purposes than knowledge of history, literature and the arts.

Nevertheless, knowledge is central to rational reflection. Having knowledge is not, of course, a matter of having mere encyclopaedic information. It is more a matter of understanding relationships between things, of seeing how facts fit into various theories, of adjudicating competently between alternative explanations of phenomena, and so forth. Someone might know, for example, as a mere matter of fact, that the Catholic church

does not permit divorced persons to remarry. But if he can relate this fact to other aspects of Catholic doctrine, to views about the nature of religious belief, to the sociobiological needs of human beings, and to other matters, clearly he has much more to reflect about than someone whose knowledge is at the merely informational level.

We have said that what is needed for rational reflection is something to reflect about, i.e. knowledge, and also something to reflect with. That is, the reflective person needs not only various kinds of knowledge that (not just in the informational sense but in the sense that contributes to deeper understanding), but also the skills involved in knowing how to think. It is sometimes thought that there are various skills involved in learning how to think, which can be acquired and applied in any area, e.g. in thinking historically, scientifically, philosophically, etc. But that is not so. It is an empirical question to what extent "effective" thinking in one area transfers to another, but there are non-empirical considerations indicating that thinking skills cannot be learned in isolation and applied across the board when reflection is required. Our criteria for determining when someone is thinking effectively, or has mastered certain thinking skills, exhibit themselves in relation to thought content. We can tell that someone has skills to think with only in relation to his having something to think about. We recognize achievement in some field of knowledge as a manifestation of good thinking skills. Thus neither the content nor the skills of rational reflection can be specified

in advance. Whether one reflects well or poorly, broadly or narrowly, deeply or superficially, in any given instance depends on what considerations are relevant in that particular case.

One kind of consideration that is always relevant, however, is morality. So it is worth discussing briefly the relationship between autonomy and moral agency. Two questions arise: 1) Is the reflective sociopath autonomous? and 2) What is the difference between the reflective sociopath and the autonomous, immoral agent? The sociopath, of course, is not moved by ordinary moral considerations. Moral considerations do not function for him as reasons. The difference between our assessments of the sociopath and the immoral agent is that we judge the former to be incapable of being moved by moral considerations, but not the latter. Sometimes an immoral agent may appear to function very much like a sociopath, expressing little or no guilt or remorse as a result of his wrongdoing. In this kind of immoral agent, however, there is usually a psychological explanation in terms of his splitting good from bad within himself and projecting onto others the bad parts, leaving nothing in himself to feel remorse about. The sociopath, however, does not split and project in this way. He simply is not moved by moral considerations. He has no sense of badness to project. Since he is lacking in reflective ability in a very important way, his autonomy is limited, and we do not regard him as capable of moral agency. Although we may lock him up to protect ourselves from him, we do not see him as an appropriate object of retributive feelings.

Normally, we regard the agent who, like the sociopath, feels no guilt or remorse for his wrong doing, but who, unlike the sociopath, rids himself of guilt by projecting it on to others, as at least capable or moral agency, and hence as having greater capacity for self-construction than the sociopath. The agent who acts wrongly but feels guilt and remorse afterwards, is even more capable of autonomy (however lacking in self-control or however unreflective he may be) because he is less dominated by unconscious wishes than the "projecting" immoral agent, and unlike the sociopath he is capable of being moved by moral considerations. The point of comparing the amoralist with the immoralist is to bring out the importance, for rational reflection, of moral considerations functioning as reasons. Although we must allow for individual differences with regard to what is to count as a reason, a consensus on certain basics must be assumed. As Nagel points out, if an agent were to claim (sincerely) that the pain caused to someone by sticking a pin into him gives the agent no reason for not sticking in the pin, we would judge the agent to be making a mistake (1980, p.114). Relating this to the agent's ability to engage in rational reflection, we would have to conclude that it is seriously impaired.

Apart from the connection between autonomy and a subject's candidature for moral agency, there is a further connection between autonomy and moral agency concerning the relation between increased autonomy and the principle of respect for persons. It is this. One of the main concerns of psychotherapy

is that a subject should have realistic perceptions about the importance of his own needs, wishes, hopes, fears, aims, and so forth, in relation to those of other people. The two main shortcomings in this context are grandiosity and idealization. Roughly speaking, grandiosity is seeing one's own aims, goals, purposes, etc. as disproportionately more significant or important than those of others. Idealization is seeing one's own aims, goals, purposes, etc. as disproportionately less significant or important than those of others. The principle of respect for persons requires that others be seen realistically as ends in themselves. Grandiosity and idealization are in opposition to the principle of respect for persons because they treat others as mere means to satisfying certain psychological needs on the part of the agent. Thus there is a second ground for emphasizing a connection between autonomy and moral sensitivity (cf. Toulmin, 1977, pp.314-316). The point of the previous paragraph was that moral sensitivity is required for autonomy. The point made here is that autonomy is required (to some degree) for moral sensitivity.

Since reflection is something one can do well or poorly, the question also arises as to whether ignorance and low intelligence prevent agents from acquiring full autonomy. Telfer suggests that knowing how to think is neither a necessary nor a sufficient condition of intellectual autonomy. She says:

A piece of research by a schoolchild into some historical problem might well be autonomous, in that he refused to accept the established conclusions without questioning them for himself, and so on. But it might at the same time be full of non sequiturs, confused between historical

investigation and moral evaluation, riddled with unsupported and implausible surmise, and in general show that he does not know how to think historically. Conversely, a person may know how to think (in some particular sphere or in general) but nevertheless on some occasions, perhaps through laziness or bias, accept uncritically implausible views from others, produce half-baked arguments, and so on. I suggest then that knowing how to think is neither a necessary nor a sufficient condition of intellectual autonomy.

(1975, pp.31-31)

In spite of the fact that the schoolchild in Telfer's example does not think very well, he nevertheless does a good job of thinking for himself. He demonstrates autonomy in reflecting upon the established conclusions in an independent way. But if the same piece of research were produced by someone from whom we expect much higher standards, a tutor in history, for example, we would say that the tutor had done a very bad job of thinking for himself. The difference is that although the level of reflection in both cases is qualitatively the same, the schoolchild reflects well in relation to what we expect of schoolchildren, and the history tutor reflects poorly in relation to others of similar age, education, and so forth. Thus the rational reflection requirement of autonomy is fulfilled when someone reflects as well as can be expected of him. Autonomy ^{of a superior standard, but it guarantees reflection} does not guarantee reflection ^{of a high standard} in relation to what can be reasonably expected of the agent.

(b) Strong Evaluation

The difference between strong and weak evaluation is the difference between qualitative and non-qualitative

assessment of desires, actions, purposes, achievements and so forth (Taylor 1977, 1982) . Suppose, for example, that a career decision is contemplated between two alternatives - becoming a philosopher, say, and becoming a medical doctor. If the decision is based on weak evaluation alone, then the relevant considerations will be fairly practical and consequentialist, e.g. the availability of jobs, anticipated income, anticipated job satisfaction, hours of work, suitability of temperament, and so on. But if the decision is also based on strong evaluation, then it will be based on views about, for example, the higher worth of intellectual pursuits over the practice of administering to ailing bodies, or alternatively, the higher worth of the art of healing over what might be evaluated by some as useless playing with words. Strong evaluation is the articulation of judgements about what is noble or base, worthy or unworthy, edifying or degrading. It is the assessment of the value of things as ends in themselves, independent of their instrumental value. Usually it takes the form of a moral or aesthetic judgement. The significance things have for us is shaped by the strong evaluations we make. Taylor (1977) contends, rightly, that strong evaluation is the central feature of the self of a responsible human agent.

There are two ways in which strong evaluation is central to self-government. The first is this. If the self of a responsible human agent is constituted, in large measure, by strong evaluations, then by making strong evaluations one acquires, as it were, "more" self. This point applies also to

non-qualitative desires and beliefs (and to the various complexes of aims, goals, purposes, ideals, aspirations, etc. built upon them); but because strong evaluation draws from another whole dimension of desire and belief not open to non-qualitative evaluators, there is a sense in which a deeper self is constructed through strong evaluation than through weak evaluation.

The second way in which strong evaluation relates to self-government is by the self-construction of fundamental strong evaluations through rational reflection. Just as relevant knowledge is needed to consider the plausibility of beliefs, so knowledge is important for the reflections that result in strong evaluations. This includes not only knowledge of the many possible objects of evaluation, both at the level of human achievement in general and the level of specific human actions, but also knowledge of the many different kinds of evaluations that can be made. Hence, possession of a sophisticated set of evaluative concepts is needed, as well as a wide repertoire of feelings from which to draw in forming strong evaluations.

There is a close connection between Taylor's (1977) notion of strong evaluation and Hamlyn's (1977) idea of self-knowledge. According to Hamlyn, psychoanalysts and psychologists have not been concerned with self-knowledge proper, but only with the causes of human behaviour and the phases of human development. Self-knowledge proper, says Hamlyn, is much more than this. Sometimes only in making an important decision about one's

future (in choosing a career, for example, or a marriage partner) does an agent come to have self-knowledge, Hamlyn argues. That is because, in making such decisions, past commitments are no longer seen as definitive or representative of the "real" self, while other aspects of the self with which the agent had not hitherto closely identified, assume a new dimension of importance. In choosing philosophy over medicine (or vice versa) one gives weight to some considerations over others, thereby becoming (or reaffirming) the kind of person who gives weight to these considerations over others. Evaluating the past and future in light of one's strong evaluations is, for Hamlyn, the acquisition of self-knowledge.

Taylor and Hamlyn both emphasize the agent's coming to see the objects of newly formed strong evaluations in a new light. The articulation of strong evaluation is sometimes difficult. It may involve a certain amount of grappling before the evaluations become clear. This grappling is neither wholly a matter of "discovering" what is already there, or of creating something anew. As Taylor (1977) emphasizes, it is the activity of refining the inchoate mass of raw material which initially constitutes one's deep unstructured feelings about what is important.

3. Self-Control

(a) Governing the Peripheral Self

So far we have considered the exercise and development of

autonomy as liberty of action in conjunction with self-construction of the governing self through rational reflection and strong evaluation. As we have said, in a non-autonomous person, the governing self and the fundamental self may not coincide, as when someone's actions are best explained by unconscious desires. But in the autonomous person both unconscious desires and peripheral beliefs and desires are checked by the self-constructed fundamental core. That is, the autonomous person is self-controlled. It is possible for someone to have a settled core of beliefs and desires which are the products or survivors of rational reflection, yet fail to be autonomous through failure to govern "unruly" elements of himself. The idea of governing unruly elements needs clarification. How can anyone possess unruly beliefs and desires?

We have said that the self-constructed self consists of fairly settled beliefs, desires and strong evaluations, which have been rationally reflected upon by the agent. If he is self-controlled, then the agent's self-constructed core of fundamental beliefs and desires controls his peripheral beliefs and desires, by accepting or rejecting them depending on how well they cohere with the central core. Sometimes an agent will make a conscious, reflective effort to determine the extent to which his peripheral self coheres with his fundamental beliefs and desires. For example, an agent might consider whether his fundamental rejection of a belief in the existence of God, is consistent with his peripheral desire to go to church on Sunday nevertheless. For some individuals, the persistence of a

desire to go to church on Sundays in spite of the rejection of a belief in God will constitute a lack of self-control, and for others it will not, depending on whether these two apparently inconsistent aspects of the self can be reconciled by some other aspect, such as the belief that a touch of inconsistency here and there is not a bad thing. To take another example, this time of wayward beliefs rather than of wayward desires, consider the tendency that many rational people have to believe in horoscopes or fortune-telling. If a belief in such things is in opposition to the agent's most deeply held and rationally considered beliefs, then the persistence of a belief in horoscopes or fortune-telling could be an indication that the fundamental self is failing to govern. If, however, it is part of an agent's fundamental self to hold that some apparent inconsistencies in a belief structure are permissible, or even desirable, then a belief in horoscopes and fortune-telling might be less of a manifestation of lack of self-control than it would otherwise be. Unruly desires and beliefs are those which are either logically or empirically incompatible with our most deeply held desires and beliefs. Acting on unruly desires and beliefs is often followed by the feeling of remorse or guilt - the feeling of not having been faithful to one's own deepest values.

Some individuals have greater difficulty than others in ridding themselves of beliefs and desires which do not fit their central network, even after deliberating at length and consciously deciding they do not fit. Hence it is easier for

some individuals to be autonomous than for others. Nevertheless, there are at least three ways of gaining greater self-control which an agent may employ if self-control is difficult. First, there are behaviourist techniques, which amount to various forms of classical and operant conditioning. These are commonly used, with varying degrees of success, in getting rid of phobias and undesirable habits, for example the fear of spiders or the habit of smoking. But sometimes behaviourist techniques are also used in curbing unwanted beliefs, such as when for example, sensibly or not, a psychotherapist will instruct a client to repeat silently to himself several times a days some such statement as 'The past does not determine the future!'. Second, the acquisition of psychological self knowledge of unconscious drives and motives, is often thought to provide the agent with the opportunity to find outlets for these desires which he himself approves, or for avoiding situations in which the agent knows unconscious drives are likely to take hold. Young(1980a) emphasizes the role of this kind of self-knowledge in overcoming anomie, self-deception, and neurosis. Less plausibly, he also emphasizes the role of this kind of self-knowledge in overcoming ambivalence and weakness of will, which are discussed in the following section. Third, there is the whole process of self-construction itself, including self-knowledge of the Hamlyn/Taylor kind discussed above, which facilitates self-control by helping certain aspects of the self assume a dominant position.

(b) Failures of Self-Government

Anomie constitutes a failure of self-government not because there is a failure of the central self to control unruly elements, but because there is no particular fundamental self to begin with. The term 'anomie' (meaning, literally, without law) was used by Durkheim to denote that condition of society which results from the disintegration of a commonly accepted normative code. But it is commonly used to describe a condition in which an individual is lacking an internal regulator. Most heteronomous persons have internal regulators of some sort. In having some kind of settled self, they have at least the capacity for self-control. But the anomic individual has no settled self. Hence he lacks even the ordinary potential for self-control. Young suggests that psychological self-knowledge has an important part to play in remedying the condition of anomie, in getting the anomic person to recognize for himself what produces his frustrations and resentment. Hopefully, by gaining awareness of his inner conflict he can "go on to replace the disorder with a realistic hierarchical ordering that reflects the priorities with which he would prefer to identify" (Young, 1980a, p.40). The problem with the anomic individual, however, is that there is no settled self with which he identifies and no particular self with which he has a settled preference to identify. If the anomic's lack of autonomy were a matter of self-deception or neurosis, knowledge of his "hidden" motivations may help him in gaining self-control, for with this knowledge he can take steps to make the motivation with

which he identifies effective. But the anomic person can't do that very easily, for, by definition, there are no beliefs and desires with which he fundamentally identifies. It is possible that in some cases of anomie psychological self-knowledge coupled with knowledge of the fact that this condition leads to frustration or resentment, might create a strong desire in the subject to identify with some particular settled self, in which case the anomic person might then be able to take steps towards becoming autonomous. But if he does not care about his condition to begin with, and he does not have stable, fundamental desires, it is doubtful that psychological self-knowledge could help him.

In most cases of self-deception or neurosis, however, where unconscious beliefs and desires effectively dominate conscious ones, often with disastrous consequences, knowledge of unconscious elements has a greater role to play. Unconscious beliefs and desires do not, of course, automatically disappear once they have been brought into consciousness, although bringing them into consciousness often has the effect of diminishing their hold. But if changes are to be made, knowledge of unconscious desires is, for the neurotic person, at least a necessary prerequisite. With it, he can take active steps to avoid situations in which his unconscious desires tend to dominate. Moral virtue is important here. As Young points out (1980a, p.39), it may be that in the long run autonomy is forsaken not because of the dynamics of the unconscious but for lack of courage, honesty, effort of will etc. on the part of the

agent to forge his life into a life of his own.

Hampshire (1965) has argued that the freedom of the individual consists in bringing more and more of his behaviour under intentional control. That is to say, the beliefs and desires which correctly explain the actions of a free individual are those for which the agent can supply (causally efficacious) reasons. This seems basically correct, provided the self-adopted beliefs and desires controlling the agent's actions are fundamental, self-adopted beliefs and desires. When peripheral beliefs and desires govern fundamental ones, there is as much absence of self-control as where unconscious elements take over.

It is interesting to note that Hampshire distinguishes two kinds of self-knowledge: first, the standard sort of psychological knowledge of unconscious motivations, inferred from observations of past behaviour, and second, knowledge based on awareness of conscious intention. Hamlyn (1977) denies that either of these counts as self-knowledge. As we have seen, he offers an account of self-knowledge in terms of strong evaluation, which an agent frequently attains through making important decisions about his future. All three kinds of self-knowledge are relevant to our discussion of autonomy, although it is of little import which of the three might properly be described as "true" self-knowledge. Psychological self-knowledge, i.e., knowledge of unconscious motivations inferred from observation of past behaviour, is important in the attainment of self-control, more so for the neurotic or self-deceived person than for others. Hampshire's second

account, based on awareness of conscious intentions is prerequisite to the idea of self-construction. And Hamlyn's view of self-knowledge in terms of strong evaluation is also an important part of self-construction. Probably these three, taken altogether, constitute all the self-knowledge it is possible to have.

Unlike self-deception and neurosis, weakness of will stems from a failure of the fundamental self to govern (consciously held) peripheral beliefs and desires, particularly desires. The familiar experience of weakness of will manifests itself not only in the moral realm, but in the prudential realm as well. It is often contended that there is no real phenomenon of weakness of will, in spite of the feeling which we sometimes experience and describe in this way. If that position is correct weakness of will is no threat to autonomy, for weakness of will does not exist; but there are ways of characterizing the phenomenon of weakness of will that make its existence quite plausible. Someone may fail to act as he believes he should because he fails to repress some opposing desire. If he later feels a sense of guilt, remorse, or shame, that is good evidence he did not act on his fundamental beliefs and desires. If there is weakness of will in this sense, it is clear that some persons are more prone to it than others, perhaps as a result of childhood conditioning, chemical or psychological addictions, or various other causes. Whatever the causes, it seems clear that a lack of psychological self-knowledge is not the cause of weakness of will. Thus it is not plausible that an increase in

self-awareness could do much to help the weak-willed person become autonomous. But self-knowledge in the philosophical (Hamlyn's) sense has an important role to play here. Where weakness of will is a frequent problem, the agent might ask himself to what extent he really does identify with the beliefs and desires he claims are central to him. This is where strong evaluation is relevant. A different way of evaluating himself and his actions could result in significant changes in an agent's fundamental self, causing it to coincide with his governing self. This could ensure that weakness of will no longer occurs, without the agent's necessarily changing his behaviour.

Sometimes a case of apparent weakness of will turns out to be a case of genuine ambivalence. For example, if A claims that he wants not to smoke, that he believes smoking is bad for him, and that it would be better for him and his family if he did not smoke etc., but he continues to inhale two packages of cigarettes a day, we might assume that the gap between his professed beliefs and desires, and his behaviour, is less a matter of weakness of will than genuine ambivalence. Although this smoker says he believes that smoking is qualitatively undesirable, if he continues to smoke two packages of cigarettes a day without much guilt or remorse, he probably also believes that smoking provides him with some benefit. This suggests that the problem of ambivalence is the problem of a split in the fundamental self, and therefore not a problem of self-control, but a problem of self-construction. If this two-packages-a-day,

no-guilt smoker were to reassess his beliefs about smoking, he might find that, all things considered, he does not believe smoking to be such a bad thing after all. Without changing his behaviour, he might re-construct himself in such a way that his weakness of will disappears, thereby acquiring a new evaluation of, in this case, smoking, as well as a new evaluation of his past and future behaviour. This, in Hamlyn's sense, is self-knowledge.

It should be clear that if any part of the agent other than his self-constructed fundamental core effectively governs him, the agent is lacking in self-control. It is important, therefore, that the agent wholly identify with his self-constructed fundamental beliefs and desires. Dworkin (1970) and Frankfurt (1971) suggest that identification consists in having a higher order desire (meta-desire) that some lower order desire be effective. Others have argued that the notion of identification is independent of the notion of order of desire. Watson (1975), for example, argues that we identify with courses of action, not desires, and Young (1980a) argues that where a first and second order desire conflict, identification frequently goes with the lower order desire. In answer to Watson, Glover (1985) points out that even if what we primarily value is a course of action, this will generate a higher order volition. A similar response might be made to Young: even if we identify, when there is conflict between first and second-order desires, with the first-order desire, this identification will create a second-order desire, which in turn conflicts with the other

second-order desire. When conflicts occur between second-order desires, identification, it could be suggested, consists in the formation of a third-order desire to adjudicate between them. However, there is no reason to think there might not be conflicts at the third-order level as well. The question of identification cannot be settled by moving up a level yet again, for in principle there would be no end to this procedure. Besides, we cannot clearly conceive of what a fourth, or higher-order desire would be like. The conclusion is that the notion of higher-order desire is helpful in characterizing the notion of identification only in the absence of conflict between higher-order desires. But the notion of identification need not be otherwise baffling. Suffice to say that an agent identifies with his self-constructed fundamental beliefs and desires when they reflect his deepest values, which they will if his reflection is adequate and his strong evaluations correctly articulate his unstructured sense of what is important.

(c) Self-Creation

So far we have discussed self-construction as control of the fundamental self (beliefs, desires, and strong evaluations) through rational reflection, and we have discussed self-control as both the subordination of unconscious beliefs and desires to conscious ones, and the achievement of consistency (often itself attained through rational reflection) between peripheral beliefs and fundamental ones. Lack of coincidence between the fundamental self and the governing self is sometimes a failure

of self-construction, as in extreme cases of weakness of will, and sometimes it is a failure of self-control, as in genuine cases of occasional weakness of will, or domination by unconscious desires.

There are two ways in which an autonomous individual may acquire either fundamental or peripheral beliefs and desires. First, he might passively discover their presence within himself, reflect upon them, and retain or reject them according to his reflections on their acceptability. Alternatively, there is the prospect of self-creation. Someone might take steps to actively induce various beliefs and desires in himself at either a fundamental or peripheral level. Someone might, for example, take steps to induce in himself an appreciation of modern poetry, which he has hitherto abhorred. Or he might try to cultivate in himself greater patience, courage, tolerance or integrity. If his religious faith is foundering, he might take steps to induce in himself a stronger belief in God. In all of these cases the agent takes steps beforehand in determining the content, at least to some extent, of his future beliefs and desires. Of course he cannot do this without some prior commitments. In order for him to choose to induce certain beliefs and desires in himself, already he must be committed to certain other beliefs and desires, as a result of which the newly chosen beliefs and desires are thought to be important. These prior beliefs and desires may or may not be very well reflected upon. Similarly, the project itself may or may not be very well reflected upon. To the extent that reflection is lacking

self-creation has little to do with autonomy. The more self-creation is tied up with an agent's reflections, however, the more it depicts a way of being actively self-constructed and self-controlled, as opposed to the more mundane way of being reflectively, but passively, autonomous.

Young (1980b) distinguishes between autonomy in the global and occurrent senses. The occurrent sense of autonomy relates to a subject's acting autonomously in particular situations. The global sense concerns a person's life as a whole. Relating global autonomy to the fundamental self, it seems clear that global autonomy suggests a hierarchial structuring of beliefs and desires, with global desires having not only logical priority over the rest but also a more fundamental position in the whole network of desires. Both global and occurrent autonomy are consistent with the subject's being either actively self-creative or passively autonomous; but in general, autonomy suggests a certain amount of global self-creation for the following reason.

It is possible, as we have said, that someone might be very self-creative, but not very autonomous, in the absence of reflection about what he is doing, but the reverse is not likely to occur. It is doubtful that a self-constructed person would fail to take active steps to shape his life, character and his dispositions, because part of what it is to be deeply reflective is to be a strong evaluator, which normally includes having a strong sense of which pursuits and traits of character are desirable. However, it is also possible for someone to be

very highly self-creative, as well as self-constructed and self-controlled, so that his project of self-creation dominates the whole of his life. We might tend to regard such a person as more autonomous than his self-constructed, self-controlled, but less self-creative counterpart. However, it makes no difference for the purposes of this thesis whether we view self-creation, in addition to self-construction and self-control, as a further development of autonomy (when it occurs in conjunction with these two), because as long as there is no interference with an agent's liberty, or his achievement of self-construction and self-control, it follows that there will be no interference with any project of self-creation he might have (cf. Glover, "Self-Creation," forthcoming).

B. Objections

1. Spontaneity

It is sometimes thought that, contrary to the foregoing account of autonomy, there is a close connection between autonomy and spontaneity. On Kant's view the autonomous will is spontaneous in the sense that it is uncaused, although Kant does not divorce autonomy from reason as does Sartre. B.F. Skinner attempts to deny that we are autonomous by denying the existence of homunculi. He says:

Unable to understand how or why the person we are behaves as he does, we attribute his behaviour to a person we cannot see...The function of the inner man is to provide an explanation which will not be explained in turn...he is a centre from which behaviour emanates...we say that he is autonomous.
(1972, p.14)

This passage illustrates the supposed connection between autonomous choice and absolute spontaneity. Many writers have attacked this supposition on the ground that it renders the notion of choice unintelligible. For example, in his discussion of ^{the} existentialist conception of autonomy, Dearden points out:

The notion of 'choice' would seem to require as a condition of its intelligible application that there should be criteria by reference to which the choice is made, and which are not, for the moment at least, brought into question.
(1972, p.67)

Where there is absolute spontaneity, there are no criteria by reference to which a choice is made. Taylor argues similarly:

A choice utterly unrelated to the desirability of the alternatives would not be intelligible as a choice...If we...try to make it (a choice) independent even of our de facto preferences, then we fall ultimately into a criterionless leap which can't properly be described as choice at all.
(1977, pp.122-123)

Should we dismiss the notion of a criterionless choice as unintelligible? Perhaps it is not unintelligible, but should be rejected on other grounds. Let us consider it with the help of one of Sartre's examples.

Suppose you start out on a hike with friends. Your walk for several hours and fatigue sets in. Eventually walking becomes painful. At first you try to resist the pain, but soon

you throw your knapsack down by the side of the road, refusing to go any further. As it happens, you could have resisted your fatigue much longer; you could have trudged on to the resting place with the others a mile or two down the road. Instead you give up and stop for a rest (1966, p.584-5). According to Sartre your choice to rest now instead of later is criterionless. By that he means it should be regarded as a kind of symbolic act, symbolic of your "project in being" or, less formidably, symbolic of the person you are. Sartre rejects seeing the choice of packing it in and sitting down by the side of the road as the product of antecedent causes, e.g. fatigue and a desire for rest, in favour of interpreting the experience of choice as a kind of self-expression.

Of course Sartre is well aware that a person normally weighs and considers relevant factors before making a decision like the one to sit down and rest. However, what he rejects is the view that in the process of weighing relevant considerations one is somehow led by the weightier factors, whatever they may be, to choose one way rather than the other. He rejects seeing choices in this light because he regards such a perspective as an evasion of responsibility. It portrays the chooser as a passive recipient of active, self-propelled reasonings. We are not accustomed to thinking of choices in Sartre's terms. We normally think of a choice as the outcome of weighing relevant factors, even when the only relevant factor is mere preference. Someone may choose to eat an apple rather than an orange because he prefers apples. That is a perfectly intelligible kind of

choice to make. But Sartre is suggesting we look at ordinary choices in a different way, not as the product or result of taking various considerations into account, but as symbolic of the persons we are. From Sartre's phenomenological perspective, a person simply opts one way rather than the other, at the same time opting for the reasonings that accompany the choices he makes. Acting in good faith, on this view, is taking full responsibility for our choices, rather than permitting some or all of that responsibility to fall on the reasons for choosing one way rather than the other.

Whether the notion of a criterionless choice is intelligible, seems to depend on changing what we ordinarily mean by 'choice'. Sartre holds that choice and self-consciousness are one and the same thing (1966, p.595). Can we conceive of a criterionless act of consciousness? Probably most of us can conceive of an uncaused act of consciousness, although we are not inclined to regard an uncaused act of consciousness as a choice. Of course that is because we think of choices as deliberate, but Sartre insists he is not talking about deliberate choice (Ibid. p.594).

Someone could take the view that in the above, watered-down sense, a criterionless choice is conceivable, without believing that there are in fact any criterionless choices. Alternatively, one might view any choice as if it were a criterionless act of consciousness because that is how it feels from the inside, without thereby committing oneself to adopting this view for the purposes of analytic philosophy. In this way the theory of

radical choice need not be regarded as a threat to the truth about human choice, but as one way of articulating the phenomenological experience of choosing, as it feels, or ought to feel, from the inside.

Nevertheless, even if sense can be made of the notion of a criterionless choice in terms of a causeless act of consciousness, for the purpose of elucidating the concept of autonomy that sense remains extremely uninteresting. For if we were to act with absolute spontaneity, to do so would be something like acting on impulse. It would be suddenly to find ourselves doing one thing rather than another. There would be no point in holding up such a chaotic condition as part of a character ideal.

It might be thought that this conclusion implies that an autonomous person gives up his autonomy whenever he acts spontaneously. But that is not so. It implies only that a person have reasons for what he does, and what he believes, not that he constantly think about them. An autonomous person may act on pure habit by going for a walk each afternoon at four o'clock, without each day reflecting on the benefits of fresh air and exercise. Similarly with non-habitual, spontaneous (but not in the sense of uncaused) acts. One can easily have legitimate reasons for acting without deliberating on them prior to acting spontaneously. Or one may have some overriding belief in the psychological value of frequently acting on impulse. If so, then acting spontaneously is not as uncontrolled an occurrence as it would be if there were no overriding belief in its value.

Hence the view that rational reflection is central to autonomy is immune to any objection that on this account the behaviour of autonomous people is devoid of any element of spontaneity.

2. Autonomy and the Whole of Human Freedom

It could be objected to the account of autonomy presented above that autonomy is not the whole of human freedom, and that we need, therefore, instead of a theory of autonomy-respecting paternalism, a theory of freedom-respecting paternalism. There could be two equally autonomous persons, so this objection might go, one of whom may be freer than the other because his self-constructed fundamental core of beliefs is more rational, perhaps, or "higher" in some other way. At least two responses can be made to this objection.

First, the aim of the thesis is to present a theory of justified paternalism acceptable to the liberal tradition, which makes room for various competing theories of higher and lower ways of life. Autonomy may not, on some accounts, constitute the whole of human freedom, but it is clear that its protection and promotion ensures also the protection and promotion of whatever "additional freedom" the individual may regard as important.

Second, the following point is discussed more fully in Chapter Three, but it is pertinent here too. In some circumstances there may be several competing courses of

permissible, autonomy-respecting paternalism open to a paternalist towards a recipient. For example, a parent may be equally justified as far as autonomy-respecting paternalism is concerned, in requiring a child to practise on his violin, play rugby, or take French lessons, assuming the interference not only respects the child's autonomy but also enhances his well-being. Because the imposition of each of these activities (assuming there is an imposition) turns out to be both beneficial and autonomy-respecting, there is plenty of room for parental evaluation of these pursuits as higher or lower, and for encouraging one over the others accordingly, without thereby infringing the child's autonomy rights. In this way some kinds of "positive" freedom can in fact be promoted by a liberal-minded paternalist. This point will arise again in our discussion of well-being.

A second objection along these lines to the conception of autonomy discussed in this chapter, might focus not on the freedom-giving content or higher worth of certain beliefs and desires over others, but on the size of the gap between an agent's desires, evaluations, purposes, etc., and the world in which they may or may not be fulfilled. Persons of stoical temperament, for example, may tend to construct or maintain desires and evaluations in accordance with the likelihood of their satisfaction, thereby reducing friction between themselves and the world by reducing the number of ^{their} unsatisfied desires and negative evaluations. The resilient skeptic, on the other hand, may find that his desires, aims, purposes, principles, and so

forth, are forever at odds with the people and institutions around him. It could be thought that the freer an agent happens to be, the more he will incline towards one mode rather than the other. But the account of autonomy presented in this chapter is neutral between these or other similarly opposed orientations. Certainly it is possible to regard autonomy as less than the whole of human freedom, for reasons independent of any conception of higher and lower specific content of the self. One might plausibly believe that ^{some} autonomous persons are freer in relation to the world around them than others, not because of greater liberty, self-construction, or self-control, but because of less (or more) friction between themselves and the world. If so, autonomy would be seen as less than the whole of human freedom, though it would remain a necessary prerequisite for it. But in any case, if there is more to the whole of human freedom than self-government, we are concerned in this thesis only with the latter.

3. The Argument from Socialization

The argument from socialization states that autonomy is an impossible ideal because as a matter of empirical fact our beliefs and desires are socially conditioned in ways that prevent us from acting freely or autonomously. Unlike the hard determinist view that causality logically precludes freedom, this is a factual claim about the cultural milieu in which we live.

Young (1979) argues against the argument from socialization by pointing out that someone may come to understand the social factors which cause him to have certain desires and beliefs, thereby freeing him to give them up or adopt them as his own. His aim is to defend compatibilism against the argument from socialization, and it is a conclusion of his defence that freedom is as widespread as self-knowledge, i.e. not very widespread (Ibid. p.376). But Young goes further than he has to. Self-knowledge is important for autonomy, but it is not always relevant for the freedom that is required for moral responsibility. Besides, psychological self-knowledge may help someone overcome compulsive behaviour, but equally it may not.

A better tack against the argument from socialization is to question its assumption that our "conditioned" beliefs and desires are irresistible. A pyromaniac may be motivated by a conditioned desire for stimulation of a particular kind, but it does not follow that his desire to set fires is irresistible. It is more in keeping with our ordinary intuitions that the freedom required for moral responsibility is absent, not where self-knowledge is absent, but where irresistibility occurs. Specifying what counts as irresistibility is, of course, a major perennial problem.

Whatever the correct analysis of the notion of irresistibility, however, the argument from socialization loses all traces of plausibility when we consider the various states which disqualify someone as a chooser. These are classified by

Benn as defects of practical rationality, defects of epistemic rationality, and defects of psychic continuity whether as believer or agent (1976, p.113). Defects of practical rationality include cases in which action is unaffected either by the subject's prior decisions or by a change in his beliefs, i.e. cases in which the subject seems incapable of basing his actions on anything other than immediate gratification. A kleptomaniac, for example, may decide never to steal again, but when the opportunity arises, he steals nevertheless. His prior decision makes no difference to his action. If a compulsive handwasher is reminded that he washed his hands just moments ago, he might believe his hands are clean, but this belief does not stop him from washing his hands over and over. Psychopaths are driven in a somewhat different way. They seek immediate gratification in ways that preclude their taking relevant considerations, e.g. long term disastrous consequences, into account in choosing what to do, even though they may be very much aware of such considerations. The psychopath is not weak-willed; he simply lacks the will to do otherwise.

Defects in epistemic rationality are typified by the disordered belief structure of the paranoid. He does not change his beliefs in the light of new evidence. The difference between the paranoid and others is that he inhabits a private world in which there are no public standards of correct and incorrect canons of reasoning.

Other types of defects are often more extreme than defects of practical rationality or epistemic rationality.

Some schizophrenics may, for example, conceive of themselves not as persons effecting change in the world, but as persons to whom things happen, or even perhaps as mere processes. Some individuals do not have the psychic continuity necessary for agency. Similarly, multiple and dissociated personalities show a marked lack of continuity as agents. In such cases there may be no one person to whom beliefs and actions can be attributed. Since the advocate of the argument from socialization blurs a useful distinction between those who do, and those who do not suffer from such defects as these, the onus of proof is on the defender of the argument from socialization to prove his point, not on the advocate of freedom to defend himself against it.

Besides, the argument from socialization may achieve whatever plausibility it has, by "borrowing" plausibility from the view that freedom is incompatible with causation. But practical ethics can and should be held apart from this metaphysical issue. Autonomy is in no way self-defeating because our beliefs and desires are culturally conditioned. One can believe, without contradiction, that there is no freedom at a deep metaphysical level, whilst recognizing good reasons for making distinctions between free and unfree actions and persons for practical purposes. Coherence and consistency in these practical distinctions and the judgements that follow from them, can be pursued independently of metaphysical enquiries.

CHAPTER II

THE RIGHT TO NON-INTERFERENCE

The exercise and development of autonomy is protected by various moral and legal rights to non-interference including non-interference with actions, opportunities, and bodily and mental states. The main purpose of this chapter is to establish how our moral rights to non-interference can be alienated by the right holder. This will show how a paternalistic *prima facie* violation of a right to non-interference can be a non-actual violation, and hence autonomy-respecting. Regardless of which specific rights to non-interference we have, the question of how they are alienated forms a central part of our theory of autonomy-respecting paternalism.

A secondary purpose of this chapter is to show that *prima facie* rights to non-interference belong to all who have desires, including non-human animals. This is important because it is often assumed that children, madmen, mental defectives, and other incompetents, have no *prima facie* right to non-interference in pursuing their own ends.

The various rights to non-interference are, according to Hohfeld's (1919) analysis, best described as claim-rights. Hohfeld's account is an analysis of legal rights, but it can be

applied to the moral sphere as well. Having a claim right to non-interference means that when an agent has this right, others have a duty not to interfere. That is, there is a claim against others on the part of the agent that they refrain from interfering.

There are two kinds of claim rights, positive and negative. A positive right correlates with a duty on the part of another to provide the right-holder with some particular service or benefit. A negative right correlates with a simple duty not to impede the agent. There is no discussion in this chapter of positive autonomy rights (sometimes referred to as reception rights or welfare rights), whereby A has a right to have his autonomy positively enhanced or developed by others. There may in fact be many positive autonomy rights, such as the right to receive in our society a certain quality of education; but the various rights to non-interference are negative claim rights. They correlate not with duties on the part of others to do or provide anything, but with duties to refrain from doing certain things. In what follows we shall discuss first the source of the right to non-interference, then who possesses it, how it is infringed, and finally, how it is alienated by the right holder.

A. Source

A's prima facie right to non-interference is a moral presumption in favour of non-interference on the part of others towards him.

It is not necessarily an actual right. Where does it come from?

"Since freedom is a principle," argue Benn and Weinstein (1971), "whatever interferes with it demands to be justified."

This is the traditional liberal viewpoint in which the *prima facie* right to non-interference is often thought to be grounded, although Rawls (1971), Dworkin (1977), and others ground liberty rights in the principle of equality.

We have discussed autonomy as an intrinsic good. The acceptance of autonomy as an intrinsic good provides a reason for acknowledging for all individuals capable of exercising autonomy, a *prima facie* moral right to non-interference based on the fact that part of what it is to value autonomy is to protect and promote it. As Nozick emphasizes:

Value is not merely something that exists out there (or in us); it also is something to which we are to have, when possible, a certain relationship. Values are to be brought about, maintained, saved from destruction, prized and valued (where this last is some descriptive term of psychology plus the theory of action). When no activity of ours can affect the value, value is to be contemplated and appreciated.
(1981, p.429)

If that is right, then the very least manifestation of valuing autonomy, which is an acknowledgement of a *prima facie* right to its exercise, should occur, it will be argued, wherever autonomy can be exercised, however minimally, even in non-human animals.

The thought behind this view is that if we claim a right to non-interference in our own case, then commitment to the universalizability principle, in the sense of making no moral distinctions without relevant differences, commits us to extending a *prima facie* entitlement to non-interference wherever the exercise of

some degree of autonomy occurs. However, not everyone agrees that there are no relevant differences in this regard between non-human animals, children, and informed adults. So we must next consider whether or not there are.

B. Possession

It is argued in this section that we have good reason to extend a *prima facie* right to non-interference not only to rational, informed, adult human beings, but also to children, mental defectives, and non-human animals - in short, anyone with desires. This is not a discussion about when the *prima facie* right to non-interference becomes an actual right, but of who has a *prima facie* right to non-interference to begin with.

It could be objected that there is little point in spending very much time exploring the grounding and extent of the prima facie right to non-interference, when we are ultimately more concerned about actual rights. For example, in a dispute between police and striking miners on a picket line, the knowledge that both parties have a *prima facie* right to non-interference does little to solve the problem of conflicting interests between them. The important question is whose rights take precedence over the *prima facie* rights of the other party, i.e. whose rights are actual. What is needed is some way

of weighing up the rights of disputants against each other, in order to determine whose claims have moral priority.

In response to this objection it should be emphasized that although actual rights are ultimately what we are more concerned about, the task of assigning moral precedence to some claims over others cannot be properly addressed until it is established which claims are in the running to begin with. It is particularly important for a theory of paternalism to clarify exactly who has a *prima facie* right to non-interference to begin with, because it is often assumed that babies, children, madmen, animals, the senile, and others, have no *prima facie* right to non-interference in spite of possessing certain other rights, such as the right to be fed, clothed, sheltered, not to be harmed, and so forth. If they have no *prima facie* right to non-interference, then the door to paternalistic intervention towards them is wide open, and so is the door to non-paternalistic intervention as long as the recipient is not harmed. If children are denied a *prima facie* right to non-interference, then others may be more likely to view them as the property of their parents or the state, than would otherwise be the case. The *prima facie* right to non-interference of competent adults is often sufficient to bar paternalistic intervention even when it would be insufficient to bar intervention on the grounds of preventing harm to others. So it is important to establish exactly who has a *prima facie* right to non-interference before considering whose rights have priority from an all-things-considered point of view in particular cases.

If we were discussing a political right to non-interference, the weak notion presented here of a *prima facie* moral presumption in favour of non-interference would be of little relevance.

Dworkin points out, rightly, that "if the right to liberty is to play the role cut out for it in political debate ... it must be a right in a much stronger sense" (1977, pp.268-9). The weak sense of a liberty right does no work in the political realm, for in the weak sense we would also have to concede, he adds, that citizens have a right to vanilla ice cream (1977, p.268). However, this is so only if rights are thought to follow automatically from desires, a dubious thesis on which the defender of a weak political right to liberty need not base his claim. It might be better to base the weak political right to liberty, as we have based the moral presumption in favour of non-interference, on the intrinsic value of autonomy, but in any case it is correct that strong liberty rights (for Dworkin in the sense of trumps over Utility) are the liberty rights most worth discussing in the political realm.

But that is not so in the context of constructing a theory of autonomy-respecting paternalism, where the right to non-interference is not in conflict with utility in the same way. The weaker notion of a *prima facie* right to non-interference is all that is needed here. As we shall see, any right to non-interference strong enough to block paternalistic intervention, is a *prima facie* right which is not subsequently alienated by the right holder. This presumption in favour of non-interference occurs at what is referred to in the Introduction

as the objective level of justification.

Now back to the question of how far the possession of a *prima facie* right to non-interference should be extended. First of all, it could be thought that someone might value autonomy as an intrinsic good only in himself, and that such a person would have no reason for extending a presumption in favour of non-interference to anyone other than himself. In answer to this kind of possibility, it has sometimes been argued that it is self-contradictory to use the fact that I am I to justify a claim to be exceptional, that is, a claim to special treatment and status. M. Singer (1963) argues that this is because the reason for claiming special status has too general an application. Since it is true of everyone that he is he, and since everyone can say 'I am I', it follows that every case is exceptional, and this is self-contradictory. But arguments which show that it is self-contradictory to claim special status on the grounds that I am I succeed, it seems, only if there is a prior commitment to the principle of universalizability. If someone agrees that it is true of everyone that the fact that he is he permits him to claim special status for himself, then it is true that if he claims special status on the grounds that he is he, he contradicts himself. But he need not agree that it is true of everyone that the fact that he is he permits him to claim special status for himself. He might hold that this is so only in his own case. If his claim to special status takes the form of the singular prescription 'Let me have special status (because I am I)' and not the moral judgement

'I ought to receive special treatment (because I am I)', he thereby avoids self-contradiction. As Hare has argued there is nothing necessarily inconsistent in the amoralist's position (1981, pp.182-187). However, we are not amoralists, and there is no point in our asking the question 'Who should possess the right to non-interference?' unless we are already committed to the principle of making no distinctions without relevant differences.³ Participation in morality probably enhances the participant's well-being, as Hare (1981) has argued, but whether it does or not we remain committed to the principle of no distinctions without relevant differences. Our adoption of autonomy as an intrinsic good is the impersonal adoption of this value. That is to say, it is the adoption of the view that autonomy is desirable wherever it occurs, not solely in oneself.

Besides, the psychology of someone who values autonomy in himself for its own sake, but who disvalues it or is indifferent towards it in others, would probably be somewhat foreign to us. Such a person would be indifferent in his assessment between person A who, for example, holds his religious beliefs for reasons which he himself adopted after much informed deliberation, and person B who holds the same religious beliefs wholly unreflectively. It is hard to imagine someone to whom it is important that his own religious beliefs be held autonomously, but who is wholly indifferent about this in relation to others. Similarly, such a person would be indifferent in his assessment between person C who, say, refrains from smoking in a crowded room out of consideration for others, and person D who refrains from smoking

because he is afraid of what others might think of him. Most of us would normally admire A for his self-constructed beliefs, and C for his self-control, whether or not we agree with A's religious beliefs or C's acknowledgement of the rights of non-smokers. It is hard to imagine that someone who values his own self-construction and self-control as intrinsic goods, could under normal conditions be sufficiently lacking in fellow feeling as to be wholly indifferent

about these goods in others. The possibility that someone could value his own autonomy as an intrinsic good whilst remaining wholly indifferent to the autonomy of others even when there are no other goods at stake, remains a logical possibility but a psychological improbability. In any case, such a person would have no interest in ensuring that his paternalistic interventions with others are autonomy-respecting.

It is clear, however, that many persons not in the least psychologically unusual, would readily restrict the possession of a prima facie right to non-interference on various grounds. Rationality is perhaps the most obvious candidate for a plausible morally relevant difference between beings in this context. It is possible to be committed to the principle of no distinctions without relevant differences, but to believe the prima facie right to non-interference should be restricted to rational persons. Because the restrictor advocates the exclusion of certain beings as candidates for the benefit of non-interference, he must show how the condition of irrationality or non-rationality gives grounds for excluding persons from possession of this right. He cannot rest his case on mere preference for rational persons,

as that is what the principle of 'no distinctions without relevant differences' rules out. He might argue that irrational and non-rational beings tend to harm both themselves and others more often than rational beings if they are given free reign. Therefore, he might say, irrational and non-rational beings are not proper candidates for the prima facie right to non-interference. However, there are two replies that can be made to this even without questioning the empirical supposition that irrational and non-rational beings are more likely to cause harm to themselves and others, and without requiring clarification of the rational/irrational distinction itself.

The first reply is that if causing harm to themselves or others is the worry behind the exclusion of irrational and non-rational beings from the prima facie right to non-interference, then it is not irrationality or non-rationality in itself that makes the morally relevant difference, but a high propensity to cause harm. Using the latter criterion as a relevant difference entails that all beings, including rational beings, ought to be denied the prima facie right to non-interference if they are highly likely to cause harm to themselves or others. If the restrictor would not also deny this right to rational adult persons who are likely to cause harm to themselves or others, then he is continuing to rely on a distinction which is not, by his own admission, morally relevant.

On the other hand, if he decides to exclude from the prima facie right to non-interference all and only those who are likely to harm themselves and others, he is open to a second objection,

namely that we cannot readily or non-arbitrarily separate beings into two groups - those who are likely to cause harm to themselves and others, and those who are not. Almost everyone, at some time, causes harm to himself and others, often unavoidably. The propensities of agents to cause harm to themselves and others can lie on any point on a continuum between certainty that they will cause harm and certainty that they will not. There is no non-arbitrary cut-off point along this continuum for dividing those who are likely to cause harm from those who are not. Clearly no arbitrary distinction is morally relevant to deciding how far the prima facie right to non-interference should be universalized. So the criterion of likelihood to cause harm to oneself or others is inapplicable at the prima facie level.

Besides, an agent may, for example, have an actual right to grow petunias in his garden even though his dotty neighbour hates petunias and is positively harmed by looking at them. Because not all harm-causing actions are to be actually disallowed, and because no one can have an actual right to non-interference without also possessing a prima facie right to non-interference, the criterion of 'causing harm' does not provide a satisfactory basis for denying someone a prima facie right to non-interference.

When we restrict the right to non-interference to non-harm-causing agents, as we sometimes do, this restriction should be at the level of actual (as opposed to prima facie) rights where all the various harms resulting from the actions of the participants in any particular conflict can be weighed

up against each other. It might be objected here that morally relevant distinctions at the 'actual' level often seem to be based on arbitrary cut-off points along a continuum. If arbitrary cut-off points are relevant at the 'actual' level, so the objection might go, why are they not also relevant at the prima facie level. For example, quite apart from legal considerations let us suppose a party goer (acting as a responsible moral agent) must somehow decide (at the 'actual', or all-things-considered level) whether the whisky he has drunk is sufficient (morally) to prohibit him from driving on public roads. At the lower end of the continuum (one ounce of whisky) nothing prohibits him from driving, but at the upper end (one bottle of whisky), clearly he ought not to drive. It might be thought that in deciding at which point on this continuum he is likely to have poor control of the wheel, he is forced to make an arbitrary distinction on which is based his final moral judgement about whether he ought to drive. But his final moral judgement is not necessarily arbitrary if he competently weighs up the rights of others not to suffer the risk of harm, against his own liberty rights, and both of these in the light of the estimated probability of his having diminished control at the wheel. By contrast, in attempting to restrict the prima facie right to non-interference to those less likely to cause harm, there is no comparable weighing up of conflicting claims. That restriction is necessarily arbitrary. Where there are justified restrictions among conflicting claims, there must be some weighing up of those claims in order to determine who should

be restricted; therefore, the non-arbitrariness of decisions at the 'actual' level lies. But there is no room for the weighing up of conflicting claims at the prima facie level because at that level we are trying to determine who has a claim to begin with.

The defender of rationality as a criterion for restricting the prima facie right to non-interference could take another tack. Instead of resting his case on the claim that rational persons are less likely to cause harm to themselves and others, he may argue that rationality is an intrinsic good. As an intrinsic good, it is automatically a morally relevant criterion, so it might be thought, for restricting the prima facie right to non-interference, and perhaps many other rights as well. Why, the restrictor might ask, should the intrinsic value of autonomy have priority over the intrinsic value of rationality in determining who possesses a prima facie right to non-interference. If autonomy as an intrinsic good can generate rights, so can rationality. But to this it could be answered that a certain degree of non-interference is logically required for autonomy, and this is not the case with rationality. Denial of the right to non-interference always affects the exercise or development of the recipient's autonomy, but it does not always affect his rationality. A motor cyclist, required by law to wear a helmet, has his autonomy restricted (however minimally) by this law, but his rationality is unaffected. If the intrinsic value of autonomy can generate a prima facie right to non-interference, then no doubt the intrinsic value of rationality can generate certain rights as well; but they should

be rights intimately connected with the good of rationality itself, not some other good.

It could be argued, of course, plausibly, that a coherent and consistent network of moral thinking could revolve around the criterion of rationality as a morally relevant difference between people, and that this network might acknowledge a prima facie right to non-interference for rational or potentially rational persons only. However, it cannot follow from the coherence and consistency alone of a moral viewpoint that non-interference ought to belong only to rational or potentially rational persons, for perhaps other coherent and consistent systems could be produced based on intelligence, say, or good health, or various kinds of accomplishment. Coherence and consistency alone are not enough.

In spite of the above arguments, it remains a viable option, nevertheless, to restrict the prima facie right to non-interference to rational or potentially rational persons, on the grounds that autonomy is intrinsically valuable not wherever it manifests itself, but only where it manifests itself in rational or potentially rational persons. This is a different starting point from taking autonomy as the basic value as far as non-interference is concerned, and bringing in rationality as a possible morally relevant difference affecting prima facie rights. This starting point is autonomy plus rationality, as a generator of the right to non-interference. This point of view is frequently adopted. But there is a good reason for rejecting it on the grounds that it leads to unacceptable consequences. If animals, for example,

as non-rational beings capable of exercising a minimal amount of autonomy (i.e. capable of acting on their desires) are denied a prima facie right to non-interference, then there is nothing to bar someone from using chemical or psychological methods to alter the physical or mental states of animals as a means of mere entertainment, provided the animal is not harmed. Most of us would reject a theory which says it is all right to give dogs and cats mind altering drugs for the entertainment of observing their unusual effects.

But it is not just interference (without good reason) with the psychological states of animals that seems objectionable. Many people object to the lack of physical freedom of animals in zoos and other places of captivity, not necessarily because they suffer, but because they are unfree. It does not follow, of course, from granting animals a prima facie right to non-interference that zoo-keeping is wrong. That depends on whether there are any good reasons for keeping animals in zoos. However, it does follow that the freedom of animals, independent of the matter of suffering, is not wholly negligible. How much weight it should carry depends on the circumstances of particular cases.

We conclude that anyone who values autonomy as an intrinsic good - wherever it may occur, not just in himself - has a reason for acknowledging a prima facie right to non-interference wherever the exercise of any minimal degree of autonomy can occur. This includes non-human animals. If a dog has a prima facie right to non-interference, then it would be wrong to give him mind altering drugs without good reason. What constitutes a good

reason in some particular circumstance is a problem for determining whether the dog, in that dog's circumstances, has an actual right to non-interference. But that is another topic. The present point is that it follows from our adoption of autonomy as an intrinsic good in itself, that we have reason for acknowledging a prima facie right to non-interference for all beings with psychological states amounting to desires. The possession of desires, whether actual or potential, is important because desires give point to the right to non-interference in the following ways.

The right to non-interference₁, which is discussed in the next section, presupposes that the agent has desires which find expression in his actions. The right to non-interference₂ presupposes that the agent wants, or would want under certain conditions, to take advantage of various opportunities that might be open to him. Deprivation of either of these rights is a deprivation of liberty, and hence of the agent's autonomy. The connection between desires and the right to non-interference₂ is a little more complex, but no less fundamental. Non-interference₂ is non-interference with bodily and mental states. Normally, agents actively want non-interference with their bodily and mental states, so that either kind of interference usually contravenes the agent's wishes to begin with. But that consideration is not the heart of the connection between non-interference₂ and the agent's autonomy. Nor is it the contention here that agents have a prima facie right to everything they desire. Besides, there are some circumstances -

someone in a deep depression, for example - in which an agent may be indifferent, as long as he is not otherwise harmed, as to whether his bodily or mental states are interfered with. But even if the agent is himself indifferent, interference with his mental states inhibits the requirements of self-construction and self-control. Hence it is an affront to his autonomy. Self-construction and self-control could never get off the ground, of course, unless the agent has desires and beliefs of some kind. We shall see later that if the agent subsequently approves_e (technical term) of such interference, the apparent violation of his autonomy is non-actual, but that is the subject of II D below.

If an agent is indifferent to interference with his physical states, the connection between autonomy and the relevant right to non-interference is, admittedly, more tenuous than the connection between autonomy and non-interference with mental states. Interference with bodily states may not inhibit an agent's self-construction or self-control; and if the agent is himself indifferent, there may appear to be little connection between his autonomy and the right to non-interference with his physical states. It might be possible to trace out such a connection with the help of the idea of a property right, but in any case the general connection between desires, autonomy, and the various rights to non-interference is strong enough to support extending *prima facie* autonomy rights to all beings with desires.

Thus the right to non-interference does not apply to plants or inanimate objects. One could argue that trees, flowers,

the wilderness, or other entities without psychological states possess the right to non-interference on some grounds other than their capacity to exercise some minimal degree of autonomy - on the grounds that, for instance, they have an interest in not being interfered with. But it is doubtful that entities without psychological states are proper candidates for having interests, except in a metaphorical sense, and there seems to be no other plausible ground for giving trees and flowers rights. However, it does not follow from this that there is no duty not to preserve or help preserve the wilderness for the benefit of others.

C. Infringements

The distinction made by Thomson (1976, p.10) between infringing and violating a right is followed in this section⁴. An infringement of a right occurs when the correlative duty is not carried out. An unjust infringement of a right constitutes a violation of it. Thus, all infringements are prima facie violations, and unjust infringements are actual violations. When a right is justly infringed, it is overridden.

An infringement of the right to non-interference is interference with either the exercise or development of an agent's autonomy, including liberty, self-construction and self-control. The exercise or development of autonomy may be inhibited through interference with an agent's actions, mental and physical states, and opportunities. Autonomy may also be interfered with in

other, more indirect ways, such as incapacitating an agent by depriving him of food and shelter; but the right to food and shelter is based primarily on having an interest in staying alive, rather than an interest in autonomy. So infringements of reception rights such as the right to food and shelter, as well as infringements of certain powers and immunities, are relevant here only if they constitute, as they sometimes might, a direct infringement of autonomy. For example, if A is about to eat a bar of chocolate which, unknown to him, has been contaminated with rat poison by Animal Liberationist extremists, and B stops him from doing so, this may constitute a direct, though obviously justified, infringement of a liberty right.

As we have said, the main aim of this chapter is not to specify in detail exactly what rights to non-interference there are, but to show how the many rights to non-interference can be alienated by the possessor. The former task is important for practical purposes, but the latter is more important for showing how paternalism can be autonomy-respecting. Before addressing that issue, however, some specification of how the general right to non-interference breaks into rights of greater specificity may be helpful.

The right to non-interference₁, we have said, protects freedom of action, the right to non-interference₂ protects mental and physical states, and the right to non-interference₃ protects opportunities, i.e. keeps doors open. Obviously not all kinds of interference with actions, states, and opportunities are *prima facie* rights-infringements. The property line around

A's house may interfere with certain actions B might otherwise perform, but A's owning property is no prima facie infringement of his neighbour's liberty rights. Similarly with psychological states and opportunities. A's seeing an abundance of weeds in his neighbour's garden may put A into a mildly unpleasant psychological state, but this does not constitute an infringement of A's right to non-interference₂. Nor does the fact that, say, a book A wants to borrow from the library, has been borrowed by another reader, thus depriving A of a certain opportunity, constitute an infringement of his prima facie right to non-interference₃ (unless perhaps the other reader has failed to return the book on time). Infringements of a right are necessarily at least prima facie culpable infringements.

Infringements of the right to non-interference₁ are more readily identifiable than infringements of the right to non-interference₂. Physical intervention, such as pushing someone out of the way of an oncoming bus, is an infringement of the right to non-interference₁. And sometimes, so are effective threats and offers. Many writers discount offers as examples of coercion on the grounds that "when a person does something because of threats he does it unwillingly, whereas this is normally not the case when someone does something because of offers" (Nozick, 1969, p.459). Against this it can be argued that if threats and offers are inter-translatable, as they often are, then we cannot respond to an offer willingly but to its inter-translatable threat unwillingly. As Held points out:

Other things being equal, our arrangement is not substantially more, or less, coercive if it is described as 'If you don't join the union, you cannot get the job', rather than as 'If you join the union, you can get the job'. (1972, p.56)

On the other hand, it might be thought that an offer is less of an infringement of an agent's right to non-interference than a threat for the following reason. A threat compels a choice between an infringement of the right to act as one chooses and an infringement of some other right, but an offer does not. If one declines an offer, it may be thought, one's rights all remain intact. However, while it is true that the right to non-interference₁ remains intact when one declines an offer, it is not necessarily the case that the right to non-interference₂ also remains intact. If drug addict A, who is trying to cure himself, is offered a heroin injection by another addict B, he might, if he is strong willed and not yet fully addicted, refuse the offer. But if the offer causes A a great deal of psychological stress, and he wavers on the brink of giving in while B repeats his offer in the hope of wearing down A's resistance, then B's offer, whether it is accepted or not, infringes and no doubt violates A's right to non-interference with his psychological states. A resisted threat may similarly constitute a violation of this right even if the threat is a bluff. The more obvious and forceful kinds of interference with psychological states are conditioning, indoctrination, and other forms of brainwashing, but threats, offers, and various subtler kinds of psychological manipulation can constitute culpable interference with psychological states, whether this interference results in some

action on the part of the recipient or not.

Since even the smallest of actions can have some psychological effect on others, it might be objected that we have no clear-cut way of deciding when a psychological effect constitutes an infringement of the right to non-interference₂ and when it does not. Eating an ice-cream in a public park, for example, may inadvertently induce some uncomfortable psychological state in a passer-by, but that would hardly constitute culpable interference. Distinguishing between infringements and non-infringements of the right to non-interference₂ is difficult, and more difficult than distinguishing between infringements and non-infringements of the right to non-interference₁. At the moment, however, we are concerned only with characterizing very briefly the general, all-things-being-equal, prima facie presumption in favour of non-interference with actions, states, and opportunities.

Infringements of the right to non-interference₃ (opportunity-rights) may deprive an agent of liberty, i.e. the ability to act as he otherwise might, or the ability to become a self-constructed or a self-controlled individual. There are many cases of paternalistic deprivation of information, for example, particularly in medical contexts, which deprive agents of the opportunity to make a fully informed choice about one course of treatment over another. Similarly, deprivation of educational opportunities inhibits an agent's ability to acquire the knowledge needed for certain kinds of rational reflection, and hence it inhibits the development of his autonomy. This latter

kind of deprivation of opportunity could have the effect of diminishing autonomy to a far greater degree than the deprivation of those kinds of opportunities emphasized by, for example, civil libertarians. In any case, it is often argued that the kind of freedom achieved through self-construction, or something like it, is more important or fundamental than the freedom achieved through the protection of opportunities by civil liberties (Cooper, 1983; Taylor, 1979). Particular cases of the above three kinds of autonomy infringements are considered in Chapter Four. The point of this section has been simply to show the different kinds of autonomy rights that might be infringed and violated in cases of paternalism.

Some theorists have held the view that more choice and opportunity is always better than less, but that view is not advocated here. More choice does not always contribute towards greater autonomy or well-being, especially where the psychic costs of weighing up possible alternative courses of action overwhelm the agent, making him powerless to form settled beliefs and desires. Some individuals are more capable than others of withstanding the threat to autonomy that greater choice can make. Sometimes we welcome having our choices reduced because the extra burden of responsibility and the extra time and effort involved in making choices when there is more to choose from, ensure that having more choice is not always better than having less (Dworkin, 1982). However, even though it may sometimes be better for an agent to have his choices restricted, it is not

the case that others may limit his choices and opportunities for him, unless the agent himself alienates the relevant rights to non-interference. We now turn to the central issue of the chapter, the question of how a right to non-interference can be given up, or alienated.

D. Alienation

1. Conditions

The question is how an agent can give up a prima facie right to non-interference which is infringed at some time other than the time of alienation by the subject of that right. The notion of consent, freely given, immediately suggests itself as the vehicle by which an agent might give up certain rights, particularly the right to non-interference. There is a good reason why this should be so. Because an agent's desires, both actual and potential, are intimately connected with what constitutes an infringement of the right to non-interference to begin with, it seems appropriate that consent, which is an expression of desire, should determine the conditions under which the right is alienated. We cannot infringe or violate the opportunity rights of, for example, plants, because there are no circumstances in which plants want, have wanted, will want, or would want to have opportunities, or anything else. Singer

points out, rightly, that "the connection between having a right and having a capacity to desire whatever is the subject of the right should be clarified and defended" (1979, pp.82-83); but nevertheless it seems clear that if there are no conditions under which an entity desires or would desire anything, it is not a proper candidate for the possession of rights. (This thought is the basis of Tooley's (1973) theory of rights, though there is no reliance on that theory here.)

Because consent, by hypothesis, does not occur at the time of paternalistic interference, it is either prior or subsequent to it. It is sometimes thought that prior consent is sufficient in itself to alienate a subsequent right to non-interference. Suppose, to borrow an example from Carter (1977), S is a smoker who wants to quit smoking. He asks his friend, F, to interfere, using coercion if necessary, if the occasion ever arises when S and F are together and S attempts to purchase cigarettes. Suppose the occasion arises; F interferes with S and prevents him from buying cigarettes. Carter argues that in this case S's prior consent is sufficient to justify F's intervention. But what role does S's prior consent to interference play here? Suppose that at the time of interference S says to F:

Please disregard my earlier request to intervene. I have changed my mind about giving up cigarettes, and I would like you not to interfere with my purchase of them.

If F ignores this new request, and prevents S from buying cigarettes, S may afterwards resent F's action on the grounds that F completely ignored S's new instructions. S had changed his mind, informed F of this change, but he was ignored. S's resentment of F does not seem unjustified. If the resentment

is justified, prior consent is not in itself sufficient to alienate a right to non-interference. But if S had originally instructed F to ignore any different, subsequent instructions, resentment on the part of S after F's intervention may seem out of place. In a moment we shall look more closely at what may be occurring in this kind of case.

Surprisingly, it is also sometimes suggested that prior consent has nothing to do with paternalism. On this view prior consent to interference is a self-imposed limitation on future choices like any other limitation on future choice imposed for the purpose of preventing self-harming conduct. Lively argues that

Being tied to the mast, refusing to allow his men to follow subsequent orders, was simply the means chosen by Odysseus to prevent his harming himself. Paternalist intervention seems no more involved here than if I bought a house far distant from a pub because I feared becoming an alcoholic. (1983, p.152)

Securing protection for oneself by giving prior consent to someone for later intervention may seem analogous to taking other standard kinds of self-protective measures concerning future choices, but it does not follow that prior consent is irrelevant to the justification of intervention.

The relevance of prior consent might be explained in cases like that of S and F above, in terms of the obligation a paternalist might feel, if he has promised a recipient to intervene despite any subsequent attempt on the part of the recipient to release the paternalist from his earlier promise. As Parfit has emphasized, someone might feel committed by a promise made to an agent's earlier self, in spite of later

attempts by (successive selves of) that agent (who is assumed to have unclouded judgement) to release him from his promise (1984, p.328). However, not all cases of prior consent involve promises to earlier selves. And as Parfit also points out, "most of us believe that the person to whom we are committed can always release us" (1984, p.328). In the example of S and F above, to avoid complication let us assume that S is the same person at the time of requesting F to intervene and at the time of trying to release F from the agreement. If that is assumed, then the idea of commitment to an agent's earlier self in opposition to the wishes of his later self, plays no role in explaining why prior consent should be relevant to the justification of paternalism.

If we ignore the fact that in any given circumstance it might be very difficult for F to tell whether S's new request is indicative of a genuine change of mind or simply a case of weakness of will - the very thing S originally asked F for protection against - and if we assume that S's new request is an indication of a genuine change of mind, then clearly there seems no reason why the earlier request should be complied with at the expense of the later one. Because the effects of F's intervention occur only in the present and in the future, only S's present and future will are directly relevant in determining whether or not F should intervene.

However, rather than simply discounting prior consent as irrelevant to the justification of paternalism in cases where an agent has freely changed his mind, but not otherwise,

it is more helpful to view the role of prior consent in all cases as the role of evidence (but not necessarily conclusive evidence) that future consent will be forthcoming. Whether it counts as good evidence, depends on the circumstances peculiar to particular cases. In medical contexts, an adult's prior consent is excellent evidence that subsequent consent will be forthcoming. But with children, especially when treatment is non-standard or experimental, prior consent may be less reliable evidence of future consent. We consider specific cases of this in Chapter Four.

As for cases of prior consent in which the paternalist does feel the tug of commitment to the wishes of a recipient's earlier self, it could be argued that earlier selves have no right to bind later selves in an irrevocable way, on the grounds of preventing harm to others. But in any case, in the example above, if F feels committed to S's earlier self, he will have to make an all-things-considered judgement about whether to intervene with S or not. The intervention would then no longer be a case of pure paternalism.

If it is agreed that only consent subsequent to a prima facie violation of an individual's right to non-interference is directly relevant to the alienation of this right, a distinction must be drawn between the kind of consent that can alienate a prior right to non-interference and the kind of consent that is merely a fleeting mental act of acquiescence, compliance, permission, agreement, and so forth. Fleeting mental acts change from day to day. In a court of law, where it is sometimes

necessary to establish whether or not a victim of an alleged crime consented in advance, as in some cases of alleged rape or theft, a fleeting mental act of consent may be sufficient to render the accused less blameworthy than he might otherwise have been. But here we are not concerned with blameworthiness or legal rights, but with the objective conditions which must obtain for the right to non-interference to be alienated by its holder. It is^a further question, in any particular case, whether someone who interferes with another is blameworthy.

Any notion of consent adequate to the task of explaining how the right to non-interference can be subsequently alienated will have to include the ideas of free, informed consent and constancy over time, thereby providing a truer representation of the consenting agent than any fleeting mental act could do. And constancy over time will have to be balanced with centrality to, and consistency with, the agent's whole subsequent network of fundamental desires. At this point it seems better to drop the language of consent and talk instead of fundamental (in the sense described in Chapter One), long term, subsequent approval. Subsequent approval can then be cashed out in terms of greater fundamental preference over time that the past act of interference in question, i, should have occurred. We shall call this approval_e.

Approval_e is approval in the (extended) sense of having an overriding preference that a past act of interference, i, should have occurred. This is not a simplistic "thank you" theory of justification. The approval in question may or may

not be conscious or reflected upon, and it must be overriding. A preference that i should have occurred is overriding if it is stronger than any preference the agent might also have that i should not have occurred. The strength of an overriding preference is not, of course, a matter of its felt intensity, because many preferences are not consciously felt or in any way articulated by the agent. The reference here is not to unconscious desires or motives, but to those implicit preferences attributable to someone because they are inferred from the rest of the agent's whole network of preferences. Just as an implicit belief that, for example, Mount Everest is not covered in vanilla ice-cream, is correctly attributable to most people even when it has not been conscious, reflected upon, or articulated in any way by the subject holding the belief, desires can be similarly implicit rather than explicit.

An implicit desire that i should have occurred is not, of course, to be confused with a desire the agent would have if he were omniscient or aware of all the future consequences of i. An implicit preference is simply a preference that is inferred from the agent's whole network of preferences, particularly those that constitute his fundamental self.

If the strength of a preference that i should have occurred is not to be measured in terms of felt intensity, it must be measured in other ways. There are in fact three ways in which it can be measured: 1) in terms of its centrality in the agent's whole subsequent network of fundamental preferences,

2) in terms of its persistence through time, and 3) in terms of how well it coheres or fits in with the rest of the agent's whole subsequent network of fundamental preferences. Fundamental preferences, it will be remembered from Chapter One, are those which are most important to the agent, those with which he most closely identifies. This includes informed preferences, explicit or implicit, but it does not include informed preferences which would become actual only when satisfied unless they are in fact satisfied. We shall return to the idea of an informed preference in the discussion of well-being in the following chapter.

So far we have described subsequent approval_e as the appropriate vehicle for alienation of a prior right to non-interference on the grounds that the agent's preferences ought to feature in the alienation of this right, and that future preferences are directly relevant in a way that past preferences are not. But there is also another important reason why subsequent approval_e of prior interference makes that interference autonomy-respecting. The reason is this. By subsequently approving_e of a prior prima facie infringement of a right to non-interference (which he did not approve at the time of interference) the agent comes to identify with certain aspects of himself (beliefs, desires) which formerly were peripheral or even disowned. He also comes to disown or make peripheral that part of him which earlier opposed the interference. Coming to see things differently in this way is the result of self-construction. Just as most of us believe that we can be released from a promise if that is in

accordance with the promisee's later wishes, on the grounds that later wishes take priority over earlier ones, for similar reasons it seems plausible that an agent's fundamental wishes after paternalistic intervention can have priority over his fundamental wishes at the time of intervention. Later beliefs and desires cancel the authority, as it were, of earlier ones. And, subsequent approval_e ensures that paternalism and its consequences are in accordance with the agent's exercise of control over his fundamental self, i.e. his self-construction.

We normally think of respect for autonomy as respect for an agent's present fundamental beliefs and desires causally determining his present behaviour, but interference in accordance with an agent's subsequent approval_e is in fact more autonomy-respecting, because it accords more with the agent's fundamental governing self from the time of interference onwards. It is important that this be recognized in the context of justifying paternalism for the following four reasons. *First,*

it explains the significance of the intuition of many people that subsequent (free) approval or disapproval of past paternalism by the recipient is somehow relevant in determining whether the act of paternalistic interference should have occurred.

Second, it avoids the danger of stretching the notion of an informed desire beyond recognition. For example, it is an implausible argument that making a twelve year old child practice piano lessons against his will, is in keeping with

his present informed desires when he appears to be fully informed about the opportunity he forfeits in his wish not to practice. In trying to justify paternalistic intervention in such cases, someone relying on the notion of respect for autonomy as present informed desires governing present behaviour, is led either to stretch the idea of an informed desire beyond recognition, or to abandon the idea of respect for autonomy in this kind of case. Adoption of subsequent approval_e as the vehicle for alienating the prima facie right to non-interference provides another alternative.

Third, it avoids the following callous conclusion. If a fully informed, competent person acts so as to gravely harm himself, respect for his autonomy as present, fundamental beliefs and desires governing present behaviour, requires that he may not be prevented from doing so even when it is known that he would later change his mind and be grateful for the interference. Such cases may be rare, but some examples are discussed in Section IV A1 (a). The present point is not that an autonomy-respecting would-be paternalist is required to interfere in such cases (that would require further argument), but that all things being equal he is permitted to interfere provided subsequent approval_e is forthcoming.

Finally, it accommodates the importance we often place on consistency between paternalistic intervention and an agent's present fundamental desires, by treating those desires as part of the recipient's future preferences (from the time of interference onwards), and by treating them as evidence of what the totality of those future preferences will be. The view that subsequent approval_e is the vehicle by which a prima facie violation of the right to non-interference is made a non-actual violation, also

helps to explain the relevance of judgements about competence and incompetence on the part of the recipient in cases of paternalistic intervention. This will also be discussed in Section IV A1 (a).

2. Objections

(a) Adaptive Preference Formation

It might be thought by someone that the phenomenon of adaptive preference formation, the phenomenon of "sour grapes", is relevant to the determination of the genuineness of subsequent approval_e. Adaptive preference formation is the adjustment of aspirations to their possibilities. The psychological function of adaptive preference formation is the reduction of dissonance in the agent's network of desires. Unlike preferences which change through learning and experience, adaptive preferences are reversible. Elster (1982) argues that adaptive preferences, being non-autonomous, are not eligible for inclusion in a utilitarian account of welfare. Similarly, it could be thought that adaptive preferences, being non-autonomous, are not proper candidates as preferences that are capable of alienating a prior right to non-interference.

In fact, a subsequent-approval criterion for the justification of autonomy-respecting paternalism is considered inadequate by many writers, on the grounds that paternalistic interference may itself influence subsequent approval. To avoid this complication C. L. Ten, for example, suggests we use a Rawlsian veil of ignorance to determine what kinds of paternalistic interference are acceptable. This, he holds, ^{avoids} the danger that interference may be designed to shape preferences along the lines of the interfering party (1980, p.112). Lively also takes this danger as an objection to the future-consent

justification of paternalism. He says:

[A] consequence of intervention may well be (and even be intended to be) an alteration of attitudes such that agreement to the intervention will be forthcoming. Such an intention is clearly present in much parental control of children. I might force my child to go to school, to do his homework, to practise the piano (instead of or as well as counselling him on the desirability of doing so) partly at least in the hope that the compulsion will produce a taste for what is at present unattractive. In parallel, I suspect we know insufficient about the effects of legal enactments on popular attitudes to discount the possibility that laws breed a disposition to accept their purposes as legitimate. If this is so, then the strong consent argument is circular, for the present intervention is justified by the presumption of future consent, but the intervention may produce a disposition so to consent.
(1983, p.153)

However, it does not follow from the phenomenon of adaptive preference formation that an agent is unfree in the same way he would be unfree if subsequent approval of prior intervention were the result of brainwashing or indoctrination. Adaptive preferences may not be fully autonomous, but they may be preferences with which the agent identifies, and for which he is morally responsible. It is important that an agent's alienation of a right to non-interference be free, but it is unnecessary that it always be fully autonomous. There is, however, one important exception to this. It should be emphasized that when a prima facie violation of the right to non-interference is a deprivation of the exercise of autonomy, the agent's subsequent free approval_e alienates his right to non-interference, as we have said, but if the prima facie violation is a deprivation of the development or attainment of autonomy in the form of self-construction or self-control, then subsequent free approval_e is not sufficient, and minimally autonomous approval_e is required.

What should count as minimally autonomous approval_e must be left an open question. The thought behind it, however, is this. An agent is in no position to judge the disvalue to him of something which limits his ability to become self-constructed or self-controlled, unless he obtains some degree of self-construction and self-control sufficient to make an informed decision about it. For example, someone deprived of a basic education could grow up subsequently freely^{to} approve_e of this infringement of his right to non-interference. But unless he is autonomous enough to appreciate the opportunity he missed and approve_e of the deprivation nevertheless, his free approval_e is insufficient to alienate his prior right. So although we reject the phenomenon of adoptive preference formation as a threat to the view about the alienation of the right to non-interference adopted in this chapter, we make some concession to it with the idea that sometimes minimally autonomous approval_e (as opposed to simple free approval_e) is necessary for the alienation of this right.

(b) The Reductionist View of Personal Identity

It might be thought that the reductionist view of personal identity puts the justification of paternalism on a different footing. Parfit (1984, p.321) has suggested that autonomy does not include the right to impose great harm upon oneself for no good reason, because we ought to prevent agents from doing to their future selves what it would be wrong to do to other people. This view rests on a reductionist view of personal identity.

According to the reductionist view there is nothing to know about personal identity over and above the facts of bodily and psychological continuity, and the facts of bodily and psychological continuity are a matter of degree (Parfit, 1971). The non-reductionist holds the view that there is some non-reducible further fact of personal identity over and above the facts of bodily and psychological continuity (Swinburne, 1984).

Many writers have used the reductionist view of personal identity as a justification for paternalism on the grounds of preventing harm to others. Regan (1974), for example, argues that a motorcyclist is a different person in certain relevant respects after an accident from the person he was before it, and from this he draws conclusions about justifying paternalism on the grounds of preventing harm to others. On this view the 'others' are the recipient's future selves. Lively argues that this kind of view is not worth considering on the grounds that "if the view has any merit, it would simply prompt a reformulation of the anti-paternalist position so that the preventing of harm to future selves, rather than harm to oneself, would become insufficient warrant for intervention" (1983, p.160). But would it? If someone harms himself, knowingly and willingly, we are far more inclined to let him do so than if he attempts to do an equal amount of harm to someone else. There is no reason to think that we would be reluctant to interfere for the sake of preventing harm to others simply because the others concerned are an

agent's future selves.

In any case a reductionist view of personal identity is no threat to the proposal that the right to non-interference is alienated through subsequent approval_e. If we adopt the reductionist view, then there will be fewer cases of paternalism and more cases of preventing harm to others than we might have otherwise supposed, but that is unproblematic. There will also be a greater number of mixed cases, i.e. cases in which the justification for interference depends on both the case for justified paternalism and the case for preventing harm to others. Where the recipient's approval of paternalistic intervention subsequently increases, and his original identity decreases, then the alienation of his right to non-interference will simply depend on approval_e in relation to his original identity. If approval_e later follows some act of paternalistic intervention, and it is attributable not to the recipient but to the recipient's future self, then the intervention in question might be justified on the grounds of preventing harm to others, but it would not be justified on paternalistic grounds.

A reductionist view of personal identity may recommend more intervention than a non-reductionist view on the grounds of preventing harm to others, and it may add considerable complexity to the justification of intervention in particular cases, but it does not alter the general theory of rights-alienation (and hence of autonomy-respecting paternalism) presented here. That is because even if the reductionist

view of personal identity were uncontroversial, it would not provide any complete solution to the problem of justifying paternalism. For even on the reductionist view some kind of non-reductionist account of the justification of paternalism is needed for those cases in which future selves do not feature. Not all acts of paternalism have long term effects, and like the non-reductionist, the reductionist needs some account of the justification of paternalistically overriding a subject's autonomy-rights for that subject's benefit.

CHAPTER III

WELL - BEING

Besides respecting the autonomy of the recipient, justified paternalism also enhances the recipient's well-being. The purpose of this chapter is to show briefly how the paternalistic promotion of well-being can be accommodated within our theory of autonomy-respecting paternalism, that is, to show how far the principles of beneficence and respect for autonomy can be reconciled.

The chapter begins by outlining some reasons for adopting an account of personal well-being in terms of the satisfaction of informed, fundamental desires. The aim is not to provide or defend a detailed account of well-being as the satisfaction of informed, fundamental desires, but rather to specify some implications of this kind of account for other issues, such as the apparent conflict in certain cases between autonomy and well-being, and the relationship to justified paternalism of perfectionist considerations.

A. Alternative Accounts of Well-Being

1. Objective List Theories

Someone could believe that, for example, if philosophical knowledge is objectively worthwhile, then the pursuit of philosophical knowledge is good for others regardless of their past, present, or future desires on the matter. Thus, the existence of objective values could be regarded as a possible threat to autonomy-respecting paternalism.

Objective values are characterized in either of two ways, as external or internal values. External values are those which may be revealed by the attractiveness of certain things, but their existence is not dependent on the existence of interests or desires. Internal values, on the other hand, are closely connected with interests and desires.

The Platonic form of the Good is one example of external, objective value. Moore's non-natural properties are another, and so are Finnis's intelligibility criteria for human activity and commitment. Nagel remarks that he is not sure whether there are any external values, but he suggests that the objectifying tendency produces a strong impulse to believe that there are (1980, p.107). However, there are good reasons why we should not be concerned with external values for the purpose of giving an account of personal well-being.

First, even if we had some intuitive faculty by which we

could detect external value, if it has no relationship to our interests and desires, there is no good reason why we should care about it. "When moral truth is that pure," as Blackburn puts it, "there can be no reason for being interested in it" (1971, pp. 115-116). Second, questions about external values are questions about a metaphysical issue which ought to be of interest to a paternalist qua paternalist only after a practical account of what constitutes individual well-being has been constructed. A practical account is bound to be tied up with fairly ordinary notions of what is good for someone such as the promotion of his interests, or his happiness.

The objectification of internal values is more relevant than the existence of external values for the justification of paternalism. Internal values are intimately connected with the interests and desires of persons in general, while remaining independent of the interests and desires of anyone in particular. The threat for autonomy-respecting paternalism from the objectification of internal values is the same as from external values: if it were clear that something such as, say, philosophical knowledge is objectively worthwhile, then a paternalist might think the pursuit of philosophical knowledge is good for a recipient whether or not it is consistent with any of the recipient's interests or fundamental desires.

The current debate about the existence of objective values is a vast topic. Mackie (1977) has argued that only what must be included in the best explanatory picture of the world is real.

Anything else is a "queer entity" (1977, p.38). Objective values, argues Mackie, are queer entities. Nagel suggests that Mackie's argument from queerness indicates a misinterpretation of moral objectivity because the impersonal badness of pain, for example, is not some mysterious further property that all pains have, but just the fact that there is a reason for anyone capable of viewing the world objectively to want it to stop, whether it is his own or someone else's (1980, p.114). McDowell (1978, 1979, 1984) argues similarly that objective values provide an equally good or better explanation (non-causal) of value-belief than other explanations that might be produced. A state of affairs may, McDowell holds, simply be such as to merit a certain response. Against this, Blackburn (1980, 1984) argues from a projectivist perspective that we can accept the view that there are irreducibly normative truths without giving such truths the ontological status we give to physical objects. Without entering into this debate, or taking a stand on whether and in what sense objective values exist, here we shall simply take the view that since ethical cognitivism is to many people an attractive metaphysical position, we must try to remove the above-mentioned threat to autonomy-respecting paternalism. Our solution is to find an acceptable account of personal well-being independent of objective list theories of the good, and if possible to find room beyond that account to accommodate objective values, thereby effectively removing the threat without necessarily

rejecting cognitivist metaphysics. The mental state account of well-being, or some form of desire-satisfaction account are the obvious alternative candidates.

2. The Mental-State Account

According to the mental state account of well-being, a state of affairs is good for someone if and only if it provides him with increased satisfaction or pleasurable consciousness. The usual objection to the state of mind account is that there is no one state of mind common to all experiences that we value for their own sake. Sometimes we prefer one state of mind to another, although we might think both supply equal amounts of happiness. Defenders of the mental state account might try to get around this objection, as Sidgwick tried to do, by defining well-being in terms of desirable consciousness, with the understanding that there are many different kinds of desirable consciousness. But there is an overriding objection to this argument as well: we often disregard the states of mind produced by various states of affairs, and view only the states of affairs themselves as providing benefit. Often we prefer one state of affairs to another regardless of the state of mind it produces. Griffin makes this point using the example of Freud who, at the end of his life, refused the drugs which would have alleviated the pain of his illness on the grounds that he preferred thinking in torment to not being able to

think clearly (1981, p.53). Often there are states of affairs we value for their own sake which do not depend on any accompanying psychological state. In response to this objection it might be thought that Freud was putting something else before his well-being. But that response reduces well-being to something which the agent himself does not regard as either beneficial or important. It seems better, therefore, to define well-being in terms of the satisfaction of desires than in terms of mental states.

3. Well-Being as Desire Satisfaction

Desire theorists characterize an individual's good in terms of the satisfaction of preferences or desires (here the terms are used interchangeably), with various restrictions on the kinds of desires which may feature in the account. The desires considered relevant may be restricted to those that survive some kind of "cognitive psychotherapy", for example, or to those which affect only the subject's own life. There are various kinds of restrictions that can be made.

Brandt objects to any such account of well-being in favour of a mental-state account on the grounds that the whole notion of aiming to maximize the satisfaction of desires over a lifetime is unintelligible. He outlines two problems that purportedly make it unintelligible. First, there are many desires someone might have at time t for something to occur or

to have occurred at time $t + 1$ (1979, p.250). The idea here seems to be that in many cases there is no point in satisfying at time $t + 1$ a desire that occurs only at time t . The example Brandt uses is that of a six-year-old who decides he would like to celebrate his fiftieth birthday by taking a roller-coaster ride (1979, p.249). Because such desires are not satisfied at time t , on the desire satisfaction account of well-being they may be ignored. But by ignoring them we come close to a mental-state (or happiness theory) of well-being. This can be accepted, but the point makes little impact. To come close is not to coincide.

The second problem is that desires change over time.

Brandt elaborates:

Notice that one acquires some desires and loses others as one matures...Most notably, some desires are cyclical, in the sense that after satisfaction there is a period of no desire for a whole family of events, followed by a recovery of interest... In view of these facts what is a would-be maximizer of satisfaction of desires to do? If the other person's desires were fixed, you could identify his fixed long-term preference ordering of biographies for himself or the world, and then move him up to the highest indifference curve your resources permit. Since the desires are not fixed, you cannot pursue this programme.

(1979, p.250)

It is true that not all desires, perhaps not many desires, are fixed over time. But that need not present a major problem for a desire-satisfaction account of well-being, for, it might be argued, it is desires at the time of satisfaction that count.

Against this point Brandt argues:

Consider an objection...a convinced skeptic who has rebelled against a religious background wants, most

of his life, no priest to be called when he is about to die. But he weakens on his deathbed, and asks for a priest. Do we maximize his welfare by summoning a priest? Some would say not, in view of his past desires. The programme also ignores future regrets... I have the temerity to suggest the whole concept is unintelligible.

(1979, pp. 250-251)

But there is a simple answer to the question of whether we enhance someone's well-being by calling a priest to his deathbed when he asks for one after a lifetime of skepticism. If it is appropriate to describe his sudden change not as weakness of will but as a genuine change of mind, then it would seem that calling a priest would be in the agent's present and future interests. But if his sudden preference that a priest be summoned is weakness of will, then presumably he still has some second order preference that a priest not be called. Otherwise asking for a priest could not be correctly described as weakness of will. If the agent has a second order preference that a priest not be called, and this preference overrides his first order preference that a priest be called (that is, its satisfaction is required for the satisfaction of a greater portion of the agent's whole network of present and future fundamental preferences), then it would hardly be in his interests to call a priest. The usual motivation for featuring the satisfaction of past preferences in a programme for the enhancement of someone's well-being, may have to do with the fact that viewing a life as an integral whole is more satisfying than viewing it in terms of distinct past and future selves. However, a better reason for taking past preferences into account

is, as we said in Chapter II, that past preferences are frequently, though not always, reliable indicators of present and future preferences.

In Brandt's example of a deathbed conversion, if the conversion is genuine and not just weakness of will, the programme of desire-satisfaction maximization in this case is the same as the programme of happiness maximization. But that does not give Brandt his case, for the two programmes are not always the same. For example, if someone values, above all else, living in accordance with a certain religious ideal that emphasizes suffering and the denial of worldly happiness, then his well-being, but perhaps not his happiness would be enhanced by living in accordance with this ideal. One could argue, of course, that if someone values, above all else, living in accordance with this kind of religious ideal, then necessarily he would not be truly happy in living a worldly existence. If that is so, then in concession to Brandt, it must be admitted that it is difficult to determine what is at stake between the mental state account of well-being and the desire-satisfaction account. Nevertheless, for the reason given in Section A2 above, it is better to characterize well-being using the language of desire satisfaction rather than the language of mental states.

One problem for the desire-satisfaction account has to do with determining whether it is better to have a situation in which the highest possible proportion of an agent's desires are satisfied (in which case we are led to the unacceptable conclusion

that one need not have many desires to have a high level of well-being), or a situation in which the highest number of desires are satisfied (in which case we are led to the awkward conclusion that one person could be better off than another even though the former has many more unsatisfied desires than the latter). But there are reasons why the highest-proportion account should be rejected. First, even unsatisfied desires may be those which we prefer to keep rather than eliminate. This may have to do with the fact that normally we prefer to be informed rather than uninformed. If unsatisfied desires sometimes follow from a state of being informed, then even this is thought to be better than blissful ignorance. Thus most of us probably prefer to possess some unsatisfied informed desires rather than none, in spite of also having some unsatisfied desires we might be glad to be without.

Second, opting for the highest proportion of satisfied desires might lead to consequences which in normal conditions we could not accept as beneficial for anyone. For example, it leads to the absurd conclusion that a subject with only one desire - to be fed, say - is better off if his desire is satisfied than other persons with highly complex and sophisticated desires, most of which are satisfied but some of which are not.

It is also less likely to be autonomy-respecting in paternalistic contexts. A paternalist might be more inclined to deprive a subject of challenging opportunities if he believes it benefits the recipient to have fewer unsatisfied desires. We

conclude, therefore, that policies which increase desire satisfaction benefit the recipient even if they also increase desire dissatisfaction. (Exceptions to this could occur if the recipient has a strong, autonomous, fundamental desire to eliminate unsatisfied desires, but that and similar complications need not detain us here.)

It does not follow from this point, however, that all the desires which feature in a desire-satisfaction account of well-being have the same weight. The satisfaction of a global desire to live one's life as, say a philosopher, is not on a par, for the purposes of aggregation, with a local desire to attend a particular lecture. Global desires carry more weight. Global desires carry with them, suggests Griffin (1982, p.335) the relevant notion, in this context, of strength of desire.

On the issue of which of an agent's desires should feature in a desire-satisfaction account of well-being, there are important questions relating to both kind and content. Questions of kind concern actual and informed desires. The satisfaction of actual desires alone, will not suffice in a definition of well-being, because many of a person's actual desires are contrary to his real interests. So the idea of an informed desire must be introduced. But the notion of satisfying desires an agent would have if he were fully informed is also inadequate, because fully informed desires might be wholly divorced from the subject's own experience, and therefore not attributable to him.⁵

If a desire of any kind, informed or not, is attributable

to someone, it must be tied up with the subject's experience in some way, either directly or indirectly, either prior to or at the time of satisfaction. A direct connection between experience and desire occurs either if the desire has some felt intensity or if it has an explanatory role in the agent's behaviour. An indirect connection occurs if the desire is simply inferred from other desires. As we have said, just as it is possible to attribute to someone an inferred or tacit belief that, for example, Mount Everest is not covered in vanilla ice-cream, even when the person to whom this belief is attributed may never be aware of holding it, we can similarly attribute to people inferred or tacit desires. As for inferred or tacit informed desires, these may be attributable to someone on the basis of what we know about his second order preferences to be apprised of the facts, to be logically consistent, and so forth. However, an informed desire may be either inferred or directly connected with the subject's experience, and it may enter the subject's experience either prior to or at the time of satisfaction.

It was pointed out in Chapter Two that the fundamental desires featuring in the alienation of an agent's prior right to non-interference (approval_e) include informed desires, either explicit or inferred, but not desires which would become actual only if satisfied unless they are in fact satisfied. However, there is an important difference between informed desires which may feature in the subsequent alienation of a subject's right to non-interference, and the informed desires which may feature

in a programme of well-being enhancement for the same subject. For example, someone now ignorant of, say, Greek philosophy, might, if he knew the works of Plato and Aristotle, come to appreciate them deeply. Thus, in the context of setting out a programme of well-being enhancement for this person, an informed desire to read Greek philosophy may plausibly feature. But in the context of the alienation of a prior right to non-interference, unless this desire is realized either directly or indirectly in the agent's actual experience, it will not form part of the network of preferences from which approval_e is inferred. Some informed preferences are actual only when satisfied. As such, they are attributable to an agent in the context of considering for him a programme of well-being enhancement, but unless they are satisfied, they are not attributable to him for the purpose of inferring subsequent approval_e.

Questions about the content of relevant desires include questions about whether the following should be allowed to feature in the account: desires about what will happen after one's death, other-regarding desires, irrational desires, immoral desires, and so forth. The satisfaction of desires about what will happen after one's death, and (often) the satisfaction of other-regarding desires, may not enter the subject's experiences^{very} fully, so it may be thought better to exclude them. However, if they are strong desires, their satisfaction seems clearly relevant to making it true that the

subject lives a more successful life if they are satisfied than he does if these desires remain unsatisfied. As for irrational or immoral desires, if they constitute the agent's informed desires, then it is difficult to see on what grounds they should be excluded from the account, however much we should regret an agent's having them.

In conclusion, because an agent's informed desires can be either fundamental or peripheral, in the sense defined in Chapter One, it seems best to adopt an account of well-being as the satisfaction of informed desires, on the understanding that the satisfaction of fundamental informed desires is clearly better than the satisfaction of peripheral ones. Global desires will, of course, normally be fundamental ones.

B. Implications

1. Autonomy and Well-Being

There are some cases of paternalistic interference which enhance the recipient's well-being, but are not autonomy-respecting because the recipient subsequently fails to alienate his prior right to non-interference with respect to them. But there are two reasons, one empirical and one non-empirical, why such cases are very rare. First, as a matter of empirical fact people usually subsequently approve_e, even when it is possible

for them to disapprove_e, of interference that increases their informed, fundamental-desire satisfaction. There are some exceptions to this, however, especially when the interference contravenes some deeply held religious or moral principle. A blood transfusion administered to an unwilling Jehovah's Witness will probably have the overall result of greatly increasing his informed, fundamental desire satisfaction in the long run, if only because he lives longer, but it is possible that he may never come to approve_e of it. Interference in cases of this kind is (objectively) unjustified because it is not autonomy-respecting.

The second reason is more complex but also more interesting. We have said that subsequent approval_e is greater fundamental preference over time (inferred from the subject's whole network of subsequent fundamental preferences) that the act of paternalistic interference in question should have occurred. Therefore, unless the interference itself contravenes some deeply held religious or moral principle, large increases in informed, fundamental desire satisfaction (well-being) resulting from paternalistic interference will necessarily tend to be coincident with the alienation of the right to non-interference.

It is sometimes thought that large gains in well-being are sufficient in themselves to justify paternalistic intervention even when the cost of the gain is an overall diminution of autonomy, provided the diminution is sufficiently small. But

the argument of the previous paragraph suggests such cases do not occur. Surprisingly, a large cost to autonomy is consistent with a large paternalistic increase in well-being as in the case above of the Jehovah's Witness, but a small cost to autonomy is not consistent with this. That is because if there is subsequent disapproval_e of prior interference which is weighty enough to override the subsequent approval necessarily accompanying any large increase in informed, fundamental desire satisfaction, then the cost to autonomy is bound to be very high.

This is an important counter-intuitive conclusion. It is based on the intimate connection between subsequent disapproval and cost to autonomy. Because the alienation of the right to non-interference and thus respect for autonomy is based on the subject's subsequent approval of the interference in question, cost to autonomy is measured in terms of subsequent (fundamental) disapproval that the interference in question should have occurred. If disapproval is overriding (disapproval_e) the interference is not autonomy-respecting. If approval is overriding (approval_e) the interference is autonomy-respecting. Now, where there is a large, paternalistic increase in a subject's well-being, there is by definition also a large increase in his informed, fundamental desire satisfaction. Thus there will necessarily be a high level of subsequent approval of the interference (whether or not it is overriding). If it is

not overriding, i.e. if the subject fails to alienate his right to non-interference, then excluding the unlikely possibility of absolute neutrality, subsequent disapproval overrides. But because there is already a high level of subsequent approval necessarily accompanying a large increase in well-being, if disapproval overrides it, the level of disapproval will be higher still. Hence, a large paternalistic gain in well-being is inconsistent with a small cost to autonomy.

A large paternalistic increase in well-being could be consistent with a small cost in autonomy if our notion of well-being were bidirectional instead of future directed. A bidirectional account of well-being accommodates the agent's past, present, and future preferences. If past preferences counted for well-being but not for rights alienation (approval_e), then there could be large increases in well-being which satisfied some kind of amalgamation of the various preference rankings of the agent's earlier and later selves into some single, timeless preference ranking (as Brickner, 1980, suggests is possible). These could stand in conflict with some minor degree of overriding subsequent disapproval. But our conception of well-being fits the idea of paternalistic interference benefitting the agent as he is at present and as he will be in the future, rather than with the idea of benefitting the agent conceived as some integral whole from the time of birth to the time of death.

Because of the way well-being and approval_e are similarly

defined, therefore, no large paternalistic increase in well-being is consistent with a small cost in autonomy. Thus, the commonly held view that the justification of paternalism depends in each case on weighing against each other the conflicting equipollent considerations of liberty and a person's self-regarding harm (Lee, 1981), fails to take account of the common ground between them.

2. Basic Goods

The following points about basic goods are not, strictly speaking, implications of the approach to well-being adopted in this chapter. However, the discussion of basic goods is included in this section on implications because it serves to introduce certain problems which arise for the paternalist from the adoption of the foregoing account of well-being.

So far we have adopted a conception of well-being in terms of the satisfaction of informed, fundamental desires, without specifying any content for a list of what in fact normally satisfies these desires, that is, a list of basic goods. Because this is largely an empirical matter, philosophical argument gets little grip here. Nevertheless, this kind of empirical judgement has an important part to play in actual cases of paternalistic intervention.

Rawls' (1971) list of primary goods suggests itself as one possible list of goods which satisfy fundamental desires.

Rawls defines primary goods as things that every rational man is presumed to want, whatever else he wants. He distinguishes between social primary goods - rights and liberties, powers and opportunities, income and wealth, and self-respect - and natural primary goods such as health and vigor, intelligence and imagination (1971, p.62). Where does Rawls' list come from?

He says:

We must assume...that the list of primary goods can be accounted for by the conception of goodness as rationality in conjunction with the general facts about human wants and abilities, their characteristic phases and requirements of nurture, the Aristotelian Principle, and the necessities of social interdependence.

(1971, p.434)

This is a large and only partially reliable assumption, for there is no conclusive empirical evidence to tell us what in fact people want. Another problem with Rawls' list, as with any similar list of basic goods, is that it is bound to be incomplete. Why does Rawls' list not include knowledge, friendship, autonomy? Are these not goods most people want, at least to some extent, whatever else they want? Another problem is that although some things might be on everyone's list of primary goods, they might be there only in very general terms. Refined descriptions of, say, liberty, might make them more controversial. Both libertarians and non-libertarians want liberty, whatever else they want, but they vary on the amount and kind of liberty that each regards as desirable. Furthermore, Rawls' primary goods satisfy informed desires only behind the veil of ignorance. The point has frequently been made that

what people want most from behind the veil of ignorance, whatever else they want, does not always coincide with what they want most from the point of view of their more fully informed, actual desires. Income and wealth, for example, are not highly valued by members of many religious groups.

Because it would be better for our purposes (though not for Rawls' purposes) to adopt a list of basic goods in terms of what people want most from the point of view of their informed, fundamental desires, and because the goods that increase desire satisfaction vary from person to person and circumstance to circumstance, it is not possible here to produce a comprehensive list of basic goods. Still, Rawls' list is unobjectionable for our purposes as part of a larger general guideline. Although some people do not want, according to their informed desires, greater income and wealth, we know that most do. Similarly with Rawls' other primary goods.

Griffin's list of basic goods is, unlike Rawls', intended to be a direct account of what in fact satisfies people's informed desires most. Not surprisingly, Griffin's list includes among other goods, autonomy, accomplishment, deep personal relations, and enjoyment.⁶ There are many exceptions to these empirical generalizations, of course, which a paternalist may not ignore. Increases in these basic goods do not satisfy everyone's fundamental desires, and different people prefer to trade off these benefits against each other in different ways. For this reason it is not surprising that in many cases of

paternalism, disagreements about the justifiability of the intervention are mainly disagreements about whether the proposed benefit is an actual benefit, quite apart from any questions of autonomy.

Interestingly, some external objective list theories of human goods coincide with what we could expect to find on a list of "empirical" basic goods. Finnis (1980), for example, argues for seven basic aspects of well-being: life, knowledge, play, aesthetic experience, sociability (friendship), practical reasonableness and religion, which, he holds, present themselves to our understanding as valuable for their own sake when we reflect on the opportunities for human flourishing. His external list theory of intrinsic goods poses no threat, of course, to our account of well-being as the satisfaction of informed, fundamental desires, because his theory is a general theory of intelligibility criteria for human activity and commitment (1980, p.62) on which we can safely remain agnostic. But it is interesting to note that in spite of Finnis's insistence that his seven basic values are valuable independent of human desires, it appears to follow from two of his arguments that they constitute a list, like Griffin's, of what in fact satisfies informed desires.

One of the arguments is that the basic values are self-evidently valuable, and that this is apparent to one who "attends carefully and honestly to the relevant human possibilities" (1980, p.73). Thus it appears that one must be in some way

informed in order to apprehend these goods as goods. The second argument is that desires follow from the apprehension of goods as goods (rather than the other way around). Thus, desires find a place in Finnis's account, although it is not the same place they find in internal accounts (1980, pp.69-73 and 1983, pp.30-37). This suggests some connection between Finnis's list of basic values and informed desires. Admittedly, the connection is tenuous, for it stretches the notion of an informed desire a little out of shape by divorcing it from actual desire. So, for empirical purposes, we are not entitled to borrow very much from a list of basic goods such as Finnis's. Nevertheless, as a matter of empirical fact, these values probably tend to satisfy, in different ways and to varying degrees, many of most people's informed desires.

This last point is significant because it suggests that allegiance to some form of perfectionism or objective list theory (internal or external) of human goods need not conflict with the requirements of justified paternalism. This is a welcome conclusion because most paternalists possess some perfectionist conception of the good which it may be wrong for them to abandon in certain paternalistic dealings with others, particularly children. We turn now to a further exploration of this matter.

3. Well-Being and Perfectionism

There may appear to be some discrepancy between the view that paternalism should occur in accordance with the recipient's subsequent informed, fundamental desires, and the belief of many persons that we have duties towards others, especially children, to shape their desires along particular moral, intellectual, and aesthetic dimensions. But there is a way of reconciling these two views. Because we define personal well-being in terms of the satisfaction of informed, fundamental desires, it does not follow that we are inconsistent to hold that the satisfaction of some of these desires is better or more in keeping with human flourishing than others. Therein lies the relevance to paternalism of objective-list theories of basic goods like Finnis's.

An account of personal well-being in terms of informed, fundamental desire satisfaction can be regarded as a thin theory of the good. Certain perfectionist accounts of the good can co-exist with it because the satisfaction of some desires may be regarded as intrinsically better, higher, more worthwhile, more in keeping with human flourishing, etc. than others. This can lead to some uncertainty about both paternalistic permissions and paternalistic duties. (Unless otherwise stated the term 'well-being' in the following pages refers to the non-perfectionist account of well-being we have been discussing in this chapter.)

Just as political liberals believe it is wrong for

government to favour or impose any particular full conception of the good on society, many individuals often feel that paternalism, even towards children, is an improper imposition on the recipient of the paternalist's conception of the good. The difference between these two situations is that in the non-political realm, individuals have to reconcile the fact that they have some paternalistic duties, particularly towards children, with their uneasiness about imposing their own conceptions of the good on others. Ensuring that paternalistic intervention respects the recipient's autonomy and enhances his well-being, may not eliminate all the uneasiness the paternalist feels. For there is a further source of uneasiness. It is the problem that arises when there are several radically different courses of autonomy-respecting interference that might be open to a paternalist. A parent, for example, making a decision about private versus state school education for his children, can be helped only to a limited extent by the principle of respect for autonomy and a future-directed, prudential account of well-being. Both kinds of education may be accompanied by subsequent approval_e on the part of the recipient, and both may enhance the recipient's well-being. In this kind of case, given that both options are autonomy-respecting and beneficial to the recipient, grounds must be found for opting one way rather than the other. There are two ways of doing this.

First, a decision could be made based solely on considerations about what will most increase the recipient's

informed, fundamental desire satisfaction, in so far as that can be determined.⁷ Second, a decision could be made based on both considerations about what will increase the recipient's desire satisfaction and about what the paternalist considers to be intrinsically valuable. Let us first of all consider the acceptability of the first option, apart from the difficulties involved in accurately predicting what course of action, of those available, will in fact maximize a recipient's informed, fundamental desire satisfaction.

The major drawback of the first method of decision making is its apparent shallowness. According to this method the paternalist is required to ensure that his interference respects the recipient's autonomy and maximally enhances (from among the available options) the recipient's well-being. Thus the paternalist must abandon reasons for acting which have to do with his own deep evaluations - his own conception of what is worth pursuing. This is the same kind of charge of shallowness that is sometimes made against unrefined varieties of utilitarian thinking. But what makes it shallow? Just this. A paternalist's own conception of what has intrinsic worth is a very deep, fundamental part of himself. It is not easily given up or put aside. It seems right to us that a paternalist should be required to ensure that his interference with others respects their autonomy and enhances their well-being; but to require a paternalist wholly to ignore his own full conception of the good in favour of a policy of maximizing the desire satisfaction of

the recipient may seem wrong because it requires him to act in a way that partially denies his own deepest values. If to value something is to prize it, bring it about, celebrate its existence, and so forth, then if the paternalist is required to ignore his own conception of the good, he is required to degrade or disvalue in some way his own most deeply held values. There is a sense in which this requirement would make the principle of beneficence somewhat mean spirited and lacking in fellow feeling, because motivated by none of the paternalist's deepest evaluations about what is worthwhile in itself.

But the second method is not without complications either. By the second method someone may paternalistically promote his own conception of the good, provided his interference respects the recipient's autonomy and enhances his well-being (not necessarily maximally). The problem is this. Of all the courses of autonomy-respecting paternalistic action open to a paternalist, one course of action may enhance the recipient's well-being only a small amount while promoting the paternalist's conception of the good to a much greater degree, while some other course of action might enhance the recipient's well-being considerably more, but promote the paternalist's conception of the good hardly at all. What should the paternalist do? He is pulled in two directions. He cannot ignore his own conception of the good without succumbing to either shallowness or lack of fellow feeling, and he cannot ignore the fact that one course of action will enhance the recipient's desire satisfaction more

than the other.

For a paternalist who has no perfectionist account of the good, this problem does not arise. But of course most people do have perfectionist accounts of the good, and these do not always coincide with goods that will most enhance the informed, fundamental desire satisfaction of others. There is, unfortunately, no simple way out of this dilemma. We have suggested both that considerations about the recipient's well-being should weigh heavily in the paternalist's decision about whether and in what way he should intervene with others, and that it is undesirable for him to deny his own conception of the good altogether.

In cases where interference constitutes the promotion of a paternalist's own conception of the good in a beneficial, autonomy-respecting way, but where some other course of paternalistic interference would better promote the recipient's well-being, deciding what course of action one should take may be difficult. And for the paternalist the danger of self-deception may be great. However, it is not our purpose to develop a theory of paternalistic duties. Here we can only emphasize that a paternalist has good reason to give some weight to his own conception of the good in deciding what paternalistic course of action to take after ensuring that his interference respects the recipient's autonomy and enhances his well-being. He may, if he wishes, deny his own conception of the good in favour of adopting a policy of maximizing (from among the

available options) the autonomy-respecting well-being of others, but then he is open to the charge of either shallowness or lack of fellow feeling. Probably the most controversial cases of paternalism are straightforwardly those which involve only some question about whether the interference respects the recipient's autonomy or enhances his well-being. Nevertheless, the theoretical problem remains of how a paternalist can be expected wholly to ignore his own full conception of the good and opt for a policy of maximally increasing a recipient's fundamental desire satisfaction. A paternalist might, of course, succeed in "instilling" values which the recipient later autonomously adopts on his own, but that is different from the problem we have been discussing. Perhaps the best policy for a paternalist is to increase a recipient's informed, fundamental desire satisfaction to some specifiable minimal degree (though to what degree would be very hard to specify), while also promoting, in an autonomy-respecting way, his own conception of the good.

CHAPTER IV

PATERNALISM

The aim of this chapter is to apply the findings of the previous Chapters, particularly Chapters One and Two, to paternalism in a more concrete way than we have done so far. The first part of the chapter sets out some general implications of these findings in relation to both incompetence and consent. The second part is a discussion of particular cases of paternalism in a variety of interpersonal and institutional contexts.

So far we have been concerned with the justification of paternalism at the objective level only. We have specified those states of affairs which as a matter of objective fact constitute the autonomy-respecting enhancement by a paternalist of a recipient's well-being. However, the objective level of justification cannot in itself provide specific guidelines for directing paternalistic behaviour. We need moral and institutional rules of conduct (referred to on pp. 18-19 as belonging to the second and third levels of justification) tailored to help us achieve the specified objective conditions. We need,

for example, rules of conduct to tell us how the recipient's competence or incompetence, and how his consent or refusal of consent should affect paternalistic intervention. Normally, such rules will vary with varying circumstances. Although we cannot undertake here any full scale enquiry into the many familial, medical and legal contexts in which the nature and extent of paternalistic interference is often a contentious issue, there are many applicable points we can make.

A. General Implications

1. Incompetence and Paternalism

(a) The Relevance of Incompetence

In this section we consider the role that judgements about a recipient's competence or incompetence ought to have in the justification of paternalism. Philosophers typically regard incompetence as a necessary condition for paternalistic intervention (e.g. Childress, 1982; Feinberg, 1971; Hodson, 1977; Murphy, 1974; Rawls, 1971). But occasionally it has been considered a sufficient condition.⁸ The objection to regarding it as a sufficient condition is fairly obvious. It is a mistake to assume that if, for example, Jones is mentally ill, it is automatically permissible to override

his wishes on the handling of his financial affairs, or on the question of whether he should be committed to an institution, as if respect for his autonomy were wholly irrelevant. Jones might be incompetent in some respects, but competent in others. Or he might be incompetent some of the time, but competent at other times. Even if Jones is incompetent to handle his financial affairs, or to make an informed decision about his own committability, he still possesses a *prima facie* right to non-interference. If it is violated for his own good, it must be alienated by him subsequent to the intervention in the conventional way.⁹

The view that incompetence is not a sufficient condition but only a necessary condition of permissible paternalism is much more widely held than the view that it is sufficient. On this view we do not have the right to interfere paternalistically with a competent and fully informed person no matter what he may decide to do, even if he decides to commit suicide. Respect for the autonomy of others, it is usually thought, prohibits our interfering with the present desires of a person who is competent and fully informed. But consider the following case:

- (1) John is a twelve-year-old Jehovah's Witness. He is seriously injured and taken to hospital after being hit by an oncoming bus. While the doctor prepares to give him a blood transfusion, John informs the doctor that he is unwilling to accept this treatment. He demonstrates that he is informed about the consequences of his wishes, that he is emotionally stable, and that he is in general as competent to make this decision about his life as most adult Jehovah's Witnesses.

Setting aside legal considerations, or questions about hospital policy, it is clear that if incompetence is a necessary condition of permissible paternalistic intervention, the doctor is unjustified from a purely paternalistic point of view in giving John a blood transfusion.

Many people would object to this conclusion on the grounds that no twelve-year-old boy is competent to make this kind of life and death decision about himself. Admittedly, this might be true of most children of that age. But there is sometimes a tendency to smuggle substantive moral claims into our assessments under the guise of judgements of incompetence. In the above hypothetical case, John is as fully informed in the relevant respects, as reasonable, and as stable as many adults whose wishes we might be more inclined to honour if they were similarly placed. Such a case is neither a logical nor an empirical impossibility.

To this the defender of incompetence as a necessary condition of justified paternalism could respond, "Very well. If John is really that competent, he should be allowed to die if he wishes." But suppose the following occurs:

- (2) The doctor knows that John is about to be adopted into a loving, intelligent family who will provide him with an intellectually stimulating, non-religious environment. The doctor believes, correctly, that in this environment John's religious convictions will change over the next several years. He administers the transfusion against John's will. Ten years later John has lost nearly every trace of his old religious beliefs, and he thanks the doctor for having saved his life.

Given the circumstances and the (predictable) outcome of this case, it is difficult to maintain that the doctor should not have administered the blood transfusion. Thus it is difficult to maintain that incompetence is a necessary condition of permissible paternalism. The crucial factor is that the subject later freely changes his mind. Such cases may be rare, but clearly our theory of autonomy-respecting paternalism can accommodate them without adopting the callous conclusion that John should be left to die at the age of twelve in accordance with his wishes. It is, of course, often very difficult to predict accurately who will change his mind and who will not, but for the moment that is another matter. If incompetence were a necessary condition of permissible paternalism, stories like John's would have no chance of a morally permissible happy ending.

The following case is similar to (2) above in certain respects. Although it is highly unlikely that such a case would occur in the simple way in which it is described, it serves to illustrate a point about the significance of moral indignation.

- (3) Mary and John are, respectively, fifteen and sixteen years old. They want to marry - now. Legally, they are permitted to marry only with parental consent. They demonstrate as much understanding of marriage as many older couples whom we would judge to be fully competent to marry. They are fully aware of the possibility of later regretting their decision, but want to take the risk nevertheless. Their parents refuse consent, and to prevent the couple from running off together,

they apply intensive psychological pressure to a degree that violates, minimally, Mary's and John's right to non-interference₂. Later Mary and John attend different universities, make new friends, and their enthusiasm for each other fades. Each ends up happily married to someone else, grateful for their parents' prior interference.

Again, it could be suggested that no fifteen and sixteen-year-old couple is competent to decide to marry, but unless we use different standards of competence for adults as for adolescents this kind of absolutist position is not a plausible position to take. Some fifteen-year-olds are better informed and emotionally more mature than many twenty-five-year olds. The crucial difference between these age groups in this context is that a fifteen or sixteen-year-old person is more likely, for various reasons, later to regret having married so young. If paternalistically prohibited from marrying at a very young age, he or she may be more likely to make a better marriage later, and subsequently to be grateful for earlier intervention.

On the other hand if an agent is prevented against his wishes at age twenty-five from freely making a disastrous marriage, then even if he makes a successful marriage to someone else later, he is liable to resent the fact that he was not permitted to be the author of his own mistakes earlier. A stronger sense of indignation and moral outrage on the part of an older person can be expected if he is prevented from doing what he feels competent to decide to do in this kind of case, than from a much younger person. (This is not to suggest,

however, that an absence of indignation correlates with diminished strength of the desire for non-interference.) An adolescent would be less inclined to feel indignant or morally affronted, however hurt and disappointed he or she may be. There are two possible reasons for this: first, because most fifteen-year-olds are in fact much more incompetent than adults to make a decision to marry, those who interfere can be more readily forgiven for believing the adolescent in question is also incompetent. Second, on the reductionist view of personal identity, the idea of interference for the sake of preventing harm to others, namely future selves, gets more grip in the case of an agent aged fifteen or sixteen than in the case of someone aged twenty-five, because a greater subsequent evolution of future selves can be expected from the former than from the latter.

This fact about a lesser degree of indignation and moral outrage is significant, for where there is a greater sense of indignation, we have greater reason to expect that subsequent approval_e of interference will be withheld, even in cases where the interference proves to be to be beneficial in the long run. This helps to explain our different reactions to paternalistic interference in cases (3a) and (3b):

- (3a) Two competent young adolescents wish to marry. They prefer to disregard the possibility of future regrets, and it can be fairly accurately predicted they will make a disastrous marriage.

- (3b) Two competent persons aged thirty wish to marry. They, too, it can be fairly accurately predicted, will make a disastrous marriage, and they too prefer to disregard the possibility of future regrets.

Most people favour interference in (3a) but not (3b).

The point is not, of course, that it is morally impermissible to allow a competent couple aged fifteen and sixteen to marry, but that it may be morally permissible to prevent them, even if they are legally permitted to marry. This position is consistent with its being morally permissible both to allow a couple aged fifteen and sixteen to marry, if that is what they wish, and to prevent them in spite of their wishes. Similarly, the position is consistent with its being morally permissible both to allow a competent twelve-year-old Jehovah's Witness to die, if that is what he prefers to having a blood transfusion, and to prevent him, in spite of what he prefers. However, since the outcome in both of these examples is so much better for the participants if intervention occurs, paternalistic intervention is clearly morally better in such cases than non-intervention.

The following case is similar to the above cases of the twelve-year-old Jehovah's Witness and to the adolescents who want to marry, but it could be more controversial:

- (4) Mrs. Brown is eighty-eight years old. She lives alone in abject poverty, without friends or relatives to help her, and she no longer cares whether she lives or dies. Her home has not been cleaned for several months, and Mrs. Brown

no longer bathes or changes her clothes. The local health authority has offered her assistance on numerous occasions, but each time she has refused. She is in no way senile, deranged, or incompetent to decide that she no longer wants to live. However, given the conditions in which she lives, she will not survive long. Under Section 47 of the National Assistance Act, provision is made for the compulsory "removal" of persons at risk, competent or not. So the community physician forcibly takes Mrs. Brown from her home to a special hospital for elderly persons. There she bathes, wears attractive clothes, and has pleasant meals in the company of others. Within a short time she begins to enjoy her new surroundings. She is glad the community physician acted against her former wishes.

Given that Mrs. Brown is not suffering from senile dementia, lack of information, or any inability to make her own decisions, there is a strong case to be made for respecting her wishes as they stand prior to interference. On the other hand, since she is genuinely grateful afterwards, intervention appears to be objectively justified. Since the outcome of the intervention is highly favourable, we should conclude, as we have said, that in this kind of case where there is subsequent approval_e, it is better to intervene than not. The permissibility of the paternalistic intervention in a case like Mrs. Brown's, however, will in any actual circumstance probably be more controversial than the previous similar cases because the outcome is less predictable. We know that the wishes of older people are generally not as likely to change as much as those of children and adolescents. But the empirical problem of making accurate predictions about consequences is quite different from the prior

moral problem of deciding what kinds of consequences are relevant to the justification of paternalistic interference. Our concern at the moment is with the latter.

Our intuitions about the desirability of the outcome of intervention in cases like those above support the view that incompetence is not a necessary condition for permissible paternalistic intervention. But it could be objected that because these examples involve only children and elderly persons, they thereby somehow beg the question in favour of intervention while arbitrarily stipulating that the persons involved are fully competent. However, if it is conceded that there could be competent persons such as those described, then the stipulation is not arbitrary. It is required to pinpoint the role of judgments about incompetence in determining what a paternalist should do. Examples of competent middle-aged persons could be used instead, but because of the greater difficulty of accurately predicting a subsequent change of mind with competent middle-aged persons, they tend to sound a little strained.

In addition to the support the above examples in cases (1) to (4) offer, there are two other reasons why incompetence should not be regarded as a necessary condition for paternalistic intervention. First, as we have said, incompetence is often taken as a necessary condition of permissible paternalism because it is commonly believed that interference with the wishes of a competent person is an affront to the autonomy of that person. But as shown in Chapter Two, the requirements of

respect for autonomy are satisfied through subsequent approval_e of past paternalism. Second, the importance that we normally attach to the competent/incompetent distinction in the context of paternalism can be adequately explained in terms of evidence about the probability of subsequent approval_e of paternalistic intervention. If an agent is temporarily incompetent - unconscious, for example, confused, or lacking information - he is more likely subsequently to be grateful for paternalistic intervention than someone who is not suffering from any of these defects. A competent person usually prefers to take full responsibility for his own mistakes, although there are some after-the-fact exceptions to this, such as those discussed in cases (1) to (4). Neither incompetence nor competence provides conclusive evidence that subsequent approval_e of past paternalism will or will not be forthcoming, but it is often the best evidence we have.

Some theorists appear explicitly to oppose the view defended here that incompetence is evidence of future approval_e. They support another view the other way around. Murphy argues that "future consent can at most serve as evidence that ... the genuine criterion ... has been satisfied. Future consent cannot itself be the criterion" (1974, p.483). Childress argues similarly that paternalism is justified by a subject's defects, encumbrances, and limitations in decision making or acting, as well as the probable net harm of non-intervention. Both future consent and these considerations are morally relevant, he says,

but the latter are primary in justification while the recipient's probable future consent simply provides evidence that these criteria have been satisfied (1982, pp. 95-96, in opposition to Carter, 1977).

However, by future consent these writers do not mean approval_e. They mean future consent in the ordinary sense, and they are right that future consent cannot itself be the criterion of justification. Future consent in the ordinary sense might be forthcoming one day, and then withdrawn the next. Rather than regard such consent as evidence of present incompetence, however, it seems better to regard the evidential function of future consent at the objective level of justification as occurring in relation to subsequent approval_e. Of course Murphy and Childress are not focussing on the objective level of justification, but on justification levels two and three; and at those levels their view can be accepted. But it is still misleading even at those levels to view incompetence as necessary for permissible paternalism.

(b) Incompetence and Self-Regarding Decisions

So far we have been discussing incompetence as if it were clear how this notion should be understood. However, assessments of incompetence, both in and out of courts of law, are notoriously difficult to make. Usually it is thought that a person is incompetent to make a certain decision if he is

ignorant, compulsive, or devoid of reason with respect to the thing in question (Murphy, 1974). Let us consider how these criteria might be applied to judgements of incompetence.

i. Ignorance. Mill's example of someone unknowingly crossing an unsafe bridge is an obvious example of incompetence as ignorance. Mill contends, and most theorists would agree, that if we inform the person that the bridge is unsafe, and he freely chooses not to heed our advice, we have no right (from a purely paternalistic point of view) to detain him further. Interference with someone who wishes to risk crossing an unsafe bridge, and who is otherwise competent to decide for himself whether he wishes to take this risk, is subsequently not very likely to meet with his approval.

If incompetence is sometimes construed as ignorance in paternalistic contexts, it is important that the ignorance be relevant in some specifiable respect to the proposed paternalistic interference. It is clear, for example, that a patient's ignorance about some of the biochemical effects of a particular drug does not necessarily make him incompetent to decide whether he ought to take that drug. On the other hand, it is not so clear whether ignorance of philosophy makes a student incompetent to decide whether he ought to study philosophy, or whether a general ignorance such as that which accompanies various kinds of mental retardation, makes a retarded person incompetent to decide whether he ought to have children. Mill's example suggests that the

criterion of relevance should be the efficacy of the ignorance in causing an agent to do something harmful to himself. If the ignorance made no difference to the agent's actions, it would not be significant. However, the following example suggests that even where ignorance is efficacious in causing self-harm, and the agent is incompetent to decide or to do what is in his own best interests, it does not follow that he is incompetent to make his own self-regarding decisions.

- (5) Jones smokes two packages of cigarettes a day. He believes, falsely, that smoking will have little or no ill effect on him. If he were better informed he would stop smoking and live longer. Nevertheless, he strongly prefers to be free to make his own mistakes, including mistakes he makes based on his own ignorance. He considers it his privilege to smoke if he wishes, providing his smoking does no harm to others, and he freely accepts any health risk this may involve.

Although Jones is not competent to decide what is in his own best interests, he is nevertheless competent to make his own self-regarding decisions. Some people prefer to take responsibility for actions that stem from their own ignorance in a way that others do not. Allowance should be made for these differences in deciding whether harm-causing ignorance is relevant to an agent's competence to make his own self-regarding decisions.¹⁰

Unlike adults, young children seldom have an overriding desire to "own" their self-harming actions caused by their own ignorance. The younger the child, the less he seems capable of having that kind of overriding desire possessed by Jones in the

example above. This fact, in addition to the fact that children are more likely than adults to suffer from ignorance of the self-regarding effects of their actions, explains why incompetence as ignorance is often better evidence of subsequent approval_e for paternalistic interference with children than with adults.

The connection between competence and second-order desires of the kind possessed by Jones can be spelled out as follows. If an agent has a second-order desire to be the author of his own mistakes in spite of his ignorance, then his ignorance is irrelevant to his authorship. The ignorance is overridden by his desire to act according to his own lights, whether he is mistaken or not. The presence in someone of a second order desire of this kind, is often reliable evidence that under normal conditions subsequent approval_e of paternalistic interference would not ensue. In the example above, if Jones later came to believe that cigarettes are harmful to his health, it is doubtful that he would regard himself as having been incompetent in the past to have made his own self-regarding decision about smoking. He may regard his past self as incompetent to have determined what was in his own best interests, but that is a different matter.

ii. Compulsion. It is generally agreed that someone acts compulsively when he cannot help acting as he does, whatever that amounts to, because of internal or external forces beyond his control. In the context of paternalism

internal forces beyond the agent's control are more frequently relevant than external forces.¹¹ Obsessives, hysterics, neurotics, phobics, addicts, and so forth, are often incompetent to act in their own best interests because they cannot help acting as they do. Such mental disorders do not invariably produce compulsive reactions, although it seems clear that these illnesses frequently have this result. But even when a compulsive is incompetent to act in his own best interests, it does not follow that he is incompetent to make his own self-regarding decisions. Even a compulsive can have an overriding informed desire to fend for himself, as it were, while acting compulsively against his own best interests. Just as self-harming ignorance is evidence that subsequent approval_e of paternalistic interference will be forthcoming only in the absence of an accompanying overriding desire to be causally self-determining, the same is true of incompetence as compulsion.

As an example of this, let us consider two heroin addicts, A₁ and A₂:

- (6) A₁ and A₂ have both taken heroin for many months and are fully addicted. Neither can help acting as he does in injecting himself with this drug twice each day. Each wishes he could rid himself of the addiction, but only A₂ wants to do so independently. If his own willpower is not strong enough (which of course it is not) to keep him off the drug, A₂ freely prefers to live the rest of his life as an addict. If others were to keep him off the drug against his will and "cure" him, then in spite of the benefit he would be resentful. A₁, on the other hand, would be grateful for just this kind of interference.

Although both A_1 and A_2 act compulsively in taking heroin, only A_1 is incompetent to make a self-regarding decision to take it. However, both addicts are incompetent to do what is in their own best interests. A_2 may regret his predicament as much as A_1 , but he has a strong second-order desire either to cure himself, if he is to be cured at all, or failing that to remain addicted. This kind of example suggests that as with ignorance, compulsion in the sense of an irresistible impulse, is not sufficient to render someone incompetent to make his own self-regarding decisions if the subject has an overriding desire to be causally self-determining.

We normally distinguish between compulsion (irresistibility) and weakness of will. Weak-willed behaviour follows from resistible but unresisted desires. We are often inclined to think of the weak-willed person as morally or prudentially deficient in some way, but probably not incompetent to decide what is in his own best interests. On the other hand, frequently occurring weakness of will may suggest irresistibility, whereas occasional weakness of will does not. If Smith's weak-willed desire for a never-ending flow of gin and tonic occurs more frequently than similar desires in others, then it is not only more difficult for him to restrain himself on particular occasions, it is also more difficult for him to restrain himself all the time. For him, moderation is a greater burden than for others. It may seem plausible that under the weight of this extra burden Smith is at least partially incompetent to do what is in his own best

interests (cf. Feinberg, 1970, pp. 289-290).

But keeping in mind that incompetence functions as evidence that subsequent approval_e of paternalistic intervention will be forthcoming, it is doubtful that weakness of will functions as very good evidence in this way, not least because it is often accompanied by an overriding desire either at the time of interference or subsequent to it, to be causally self-determining. Still, there are many cases of weakness of will in which paternalism is a benefit and is subsequently approved_e. If A₁'s addiction to heroin has not yet reached the stage of irresistibility, his decision to inject himself could be a simple case of weakness of will. If A₁'s non-addicted friend, B, intervenes, forcing A₁ to undergo withdrawal and cure, A₁ may resist B at the time of intervention, but benefit enormously and subsequently approve_e of the help B gave him. But such a case suggests only what we have already concluded, that incompetence to decide what is in one's own best interests (or even to make one's own self-regarding decisions) is not a necessary condition of justified paternalism. Incompetence provides good evidence that subsequent approval of paternalism will be forthcoming, but the evidence is not conclusive, and it is not the only possible kind of evidence.

iii. Non-Rationality. Non rational beings include infants, animals, comatose patients, catatonics, and others in whom rationality is non-existent, rather than merely

deficient. Non-rationality is sometimes temporary, sometimes permanent. If it is permanent, then the test of subsequent approval_e is inapplicable (see note 9), because no overriding informed desire for causal self-determination is possible in a non-rational being. That is a crucial difference between incompetence as ignorance or compulsion on the one hand, and incompetence as non-rationality or irrationality on the other. At least temporary rationality is required for someone freely to have that kind of overriding desire.

If a non-rational state is temporary, then the possibility that the agent might subsequently disapprove_e of paternalistic interference must be taken into account. Very young children, being non-rational and hence incompetent to make their own self-regarding decisions, are sometimes the subjects of paternalistic interventions by persons who fail to take full account of the unlikelihood of subsequent approval_e of certain kinds of intervention - the paternalistic indoctrination of young children into religious beliefs is one example.

With temporarily non-rational adult incompetents, we often temper paternalistic interference to fit the known past wishes of the subject, sometimes requiring consent on his behalf by a spouse or relative in order to protect his interests if the subject is unable to give valid consent himself. This is as it should be, for past preferences are sometimes very reliable indicators (with the exception of children's preferences) of future preferences. None of this is problematic beyond the usual

problems associated with determining benefit, and determining whether subsequent approval_e will be forthcoming, because the identification of non-rational incompetents is fairly straightforward. Problems arise, however, with the identification of irrational incompetents.

iv. Irrationality. An agent may be regarded as irrational for at least three reasons:

- (1) Holding certain beliefs — which it is thought no rational person could believe, e.g. the belief that one is a snake. The point is not that the content of a belief determines its rationality or irrationality — the ground on which it is held does that — but that the content of some beliefs is so bizarre that it functions as evidence that they are either seriously inconsistent with the agent's main network of beliefs, or confused in some other way.¹²
- (2) Systematically weighing the available evidence incorrectly, as in cases of self-deception.
- (3) Failing to appreciate what is obviously relevant to forming a judgement on some question, e.g. failing to appreciate that the pain caused to someone by sticking him with a pin is at least prima facie relevant to whether one ought to do it or not (cf. Nagel, 1980).

It is not important for us here to consider such questions as whether, for example, a rational person could hold certain bizarre beliefs, because here we are concerned only with irrationality

as evidence that subsequent approval_e of paternalistic interference will be forthcoming. If incompetence were necessary for justified paternalism, it would be important to have the rational/irrational distinction very clearly defined. But since it is neither a necessary nor a sufficient condition, but significant only as evidence that subsequent approval_e will be forthcoming, it is less important to us whether the label 'irrational' is correctly applied to someone who, for example, believes that God has ordered him to cut off his ear, than is the empirical question of whether paternalistic interference with acts caused by such beliefs subsequently benefits the recipient and is followed by subsequent approval_e. At the same time, because we often have no way of making accurate predictions about an agent's future wishes, it is important particularly in courts of law, that the competent/incompetent distinction be refined as much as possible, and that as a general rule a paternalist abide by the present wishes of competent persons.

In the above discussions of ignorance, compulsion and non-rationality, it was argued that provided the condition is causally efficacious in preventing the agent from avoiding some harm or receiving some benefit, and provided the agent has no overriding desire for causal self-determination, then he can be regarded as incompetent to make his own self-regarding decisions. One would expect the same point to hold for irrationality as for other forms of incompetence, and so it does, except that both

self-harm and an overriding desire for causal self-determination are often much harder to recognize in irrational persons. Consider the following example:

- (7) P.Z., now twenty-seven years old, is an involuntary mental patient at a psychiatric hospital where he was committed after amputating his right hand. He had been previously committed for eight months after perforating his right eardrum and again for eighteen months after removing his right eye. For a variety of reasons P.Z. now wishes to leave the hospital, but his family are strongly opposed. They are already greatly in debt for the medical treatment necessitated by P.Z.'s penchant for self mutilation. Neither P.Z.'s parents nor his siblings can control his destructive outbursts at home, and they are frightened that he will turn on them as well. P.Z., on the other hand, maintains that he has been confined long enough, that he is not dangerous, and that, in any event, there is "no treatment for me in the hospital". He states repeatedly that "I believe in God and in brotherly love" and that "One man must sacrifice himself to God for the good of all men". P.Z. believes that God has selected him for sacrifice and maintains that he enucleated his right eye and amputated his right hand upon "direct orders from God", and that he and God thereby established a "covenant". He says "I would cut off my right foot if God told me to".

Adapted from a case prepared by P.B. Hoffman
in Beauchamp and Childress, 1979, pp.257-258.

Let us suppose that in this case financial considerations are not significant and the family are not afraid that P.Z. might turn on them as well as himself. The case is then one in which P.Z.'s welfare and respect for his autonomy are the only considerations. P.Z. clearly manifests an apparently overriding desire for causal self-determination, but in view of his beliefs, it is doubtful that we could regard his desire for causal self-

determination as free. It is not that we have any reason to believe that in principle such a person could not possibly have a free, overriding desire for causal self-determination, but we doubt the reliability of the evidence the patient provides that his desire is free in this case. We take his actions and his stated reasons for them as stronger evidence that his whole network of beliefs and desires is distorted than the verbal evidence he provides that causal self-determination is what he freely wants.

At the same time we may have lingering doubts about whether we are correct in denying that P.Z. could freely want non-intervention simply because his beliefs and desires are vastly different from our own. A similar kind of doubt may affect judgements about what constitutes self-harm. P.Z. clearly injures himself, but how are we to decide whether someone's injuries harm him if those injuries appear to satisfy his "informed" fundamental desires. In cases like P.Z.'s, the evidence is usually very difficult to access.

We normally do not regard an apparent desire for causal self-determination as genuine when the agent's beliefs and desires are so odd we find it hard to believe any rational person could have them, as in the case of P.Z. above, or when the agent fails to appreciate what is relevant to making certain judgements and decisions. As always, there is the problem that what is relevant to one person may not be relevant to another. In order to allow for these variations, perhaps only the person who is totally out

of touch, as Murphy (1974) reminds us, with those "basic, noncontroversial matters which must be regarded as relevant" (p.476) who should be considered irrational and incompetent. Fingarette (1972) calls this essential relevance.

The problem that what is relevant to one person may not be to another is reflected not only in what is seen as relevant, but also in the different ways in which evidence may be weighed. If someone systematically weighs the evidence incorrectly, as in cases of delusion or self-deception, it is less probable that an apparent accompanying desire for causal self-determination is genuine than in cases where less of the whole network of beliefs and desires is affected.

It should be emphasized that there is nothing intrinsic in irrationality, or indeed in any other form of incompetence, from which it follows that there is a prima facie case for paternalistic interference. Some people take pride in their irrational quirks, and would greatly resent interference with them. This point does not apply, however, when the idea of self-harm is built into the notion of irrationality. Culver and Gert (1982) severely restrict the notion of irrationality in this way. They say: "Only if a person harms himself without an adequate reason is he acting irrationally" (p.27). On this strict definition of irrationality, there is indeed a prima facie case for paternalistic interference with irrational persons. But that is an excessively restrictive definition. There is no reason why someone, whom we regard as very irrational because he

believes himself to be an insect, say, is necessarily harming himself. That depends on what kinds of actions follow from his irrational state of mind.

(c) Children

It is sometimes thought that there are morally relevant differences between children and adults, which are qualitative differences justifying paternalism with the former but not with the latter. As we have seen in Chapter Two, the *prima facie* right to non-interference does not belong to beings because of any attributes they possess other than desires, regardless of their level of rationality, linguistic or locomotor development, sexual maturity, or any other difference that might be pointed to between children and adults. Of course in any particular case of paternalism it is not the *prima facie* right to non-interference that is ultimately important, but the actual right, so the possibility of a morally relevant qualitative difference is not precluded by the fact that like adults, children too, have a *prima facie* right to non-interference.

But what plausible candidates are there for an actual morally relevant difference between children and adults? Locomotion, linguistic ability or sexual maturity will not suffice, for we do not give the well-being of adults who lack these attributes priority over their autonomy. Perhaps the competent/incompetent distinction, or some aspect of it, comes

close to characterizing a morally relevant difference between children and adults. But the acquisition of competence does not parallel the increase in chronological age. Newborn infants are never competent, but both adolescents and octogenarians are sometimes competent and sometimes not. As a matter of empirical fact children are more often incompetent to do what is in their own best interests and to make their own self-regarding decisions than adults, but that fact suggests only a quantitative difference between children and adults, not a qualitative one.

2. Consent and Paternalism

Particularly in medical contexts, consent of the recipient is often thought to justify what would otherwise be an act of unjustified interference. Consent is often considered important for a variety of reasons, but the most important of these is that it is thought to protect the recipient's autonomy.¹³ Persons who consent to some medical (or other kind of) procedure are exercising a form of causal self-determination which non-consenting persons do not exercise. Autonomy-protecting consent can be given, of course, only by competent persons. In order to protect the interests of incompetents, either presumed consent or proxy consent on the part of another person is usually required.

Accompanying most people's view that incompetence is a necessary condition of justified paternalism, is the view that with competent persons, consent is necessary before medical treatment (for example) to ensure non-violation of the recipient's autonomy. However, just as there is no reason to conclude that incompetence is a necessary condition of justified paternalism, there is no reason to conclude there can be no justified paternalistic intervention with competent persons in the absence of their consent. The cases previously discussed concerning the twelve-year-old Jehovah's Witness who tries to refuse a blood transfusion, the adolescents who want to marry, and the compulsory 'removal' of persons at risk, testify to that, especially when taken in conjunction with the argument that subsequent approval_e to past paternalism is consistent with respect for autonomy.

Nevertheless, cases in which paternalistic interference with competent persons is subsequently approved (and approved_e) by them are rare. Rarer still, is the accuracy of our ability to predict in any particular circumstance whether the case confronting us is a case of this kind. For these reasons the moral and institutional rule that the consent of competent persons should be obtained before treatment or other benefit is given, is important. That is, it is important for us to have an institutionalized procedure whereby certain moral and legal rights and obligations can be redistributed through consent.

If valid consent is given, it is by definition given

freely. Consent based on ignorance, compulsion, or irrationality is not valid consent. It is not the consent of a competent person. There are, of course, varying degrees of informed consent that may be given to treatment or other benefit, and varying degrees of ignorance or other defects of will that can be tolerated within the bounds of valid consent. At the second and third levels of justification, valid consent may be either past or future consent. Both past and future consent may be express, tacit, implied, or merely presumed. Judgements about these different kinds of consent are often difficult to make and highly controversial (Childress, 1982, Chpt.4). It is often argued, for instance, that the idea of presumed consent has frequently been misapplied in numerous cases of rape (Pateman, 1980). The greatest danger is in claiming too much. Locke's view that

every man that hath any possession or enjoyment
of any part of the dominions of any government
doth thereby give his tacit consent, and is as
far forth obliged to obedience to the laws of that
government

(Treatise of Civil Government, # 119)

is one notorious example among many of claiming too much.

B. Practical Implications

In Section B we first consider the morally relevant features of a number of ordinary cases of inter-personal

paternalism, largely in familial and medical contexts.

Secondly, legal paternalism is discussed. Certain problems arise for legal paternalism which do not feature in the inter-personal context. Finally, a brief discussion follows of education as an important and exceptional case of justified legal paternalism.

1. Inter-Personal Paternalism

(8) Reluctant to Learn

B is an eleven-year-old boy who hates practising piano lessons, and would not do so if his parents did not insist each day that he sit down at the piano and practise for forty minutes. B's parents, P, believe that the development of B's musical talents is good for him whether it is in accordance with B's wishes or not. Are they right to insist that B practice his piano lessons against his will?

It is sometimes thought that paternalistic interference is justified in order to prevent harm to the recipient, but not actively to promote the recipient's well-being. The thought behind this view, of course, is that in a liberal, pluralist society, no one has the right to impose his conception of the good on anyone else. However, if we accept that well-being is the enhancement of informed, fundamental, desire satisfaction, there is no need for a paternalist to make the distinction between preventing harm and promoting well-being at the objective level of justification. Preventing harm is preventing the frustration or non-satisfaction of informed, fundamental desires,

which can only be done by enhancing their satisfaction, that is, by promoting the recipient's well-being.

However, the distinction between preventing harm and promoting well-being is important in many institutional settings. Certain special relationships which hold between persons, especially parents and children, give rise to paternalistic duties to promote the well-being of others, above and beyond duties merely to prevent harm.

The first question to be considered in such a case as (8) is whether the interference benefits the recipient. On this P can only speculate about what B's informed desires will come to be. P should, of course, take into account what the informed desires are of others, in situations similar to B's. If P mistakenly believe that their interference with B's wishes will subsequently enhance the satisfaction of B's informed desires, then the interference is objectively unjustified, although the question of whether P are to be blamed for acting "unjustly" is another matter. P's blameworthiness will depend on the extent to which they could be reasonably expected to act differently in the light of available evidence about what others in B's situation usually come to want later on.

If we suppose that P's interference with B in fact results in increased informed desire satisfaction for B, and that the intervention in no way interferes with the future exercise or development of B's autonomy, there is still the question of whether B subsequently alienates his right to non-interference

with respect to his parents' intervention, that is, whether it will be more in accordance with B's subsequent fundamental desires that i should have occurred than that i should not have occurred. Again, P must consider the evidence. If most children in B's situation later come to approve_e of the fact that they were forced to take piano lessons against their will, then probably B will come to approve_e of this also. On the other hand, if B has older brothers and sisters who as children were subjected to a similar regimen, and who later claimed that in spite of having their musical talents developed, they do not approve of their parents having forced their will upon them in that way, perhaps B too, will grow up to value absolute non-interference over the development of his musical talents. Because B's future choices and informed desires are to some extent unpredictable, the (objective) justifiability of P's decision to interfere paternalistically is difficult to determine in advance. Philosophical argument cannot assist P in making the relevant empirical predictions.

It is doubtful that parents could gain support for a paternalistic stance in this kind of case by using the ground that an eleven-year-old child is incompetent to decide whether piano practice is in his interests; although as we have seen it does not follow from the fact that an eleven-year-old child is competent to make such a decision about himself that interference is unjustified. However, it is possible that there may be non-paternalistic grounds on which P's interference with B could be

justified. If P value the development of musical ability very highly, an argument could be made to the effect that it may cause them considerable harm to see their children grow up without developing their talents in this way. If the harm is substantial enough, then coercing B into practising piano lessons may be justified on the grounds of preventing harm to others, namely P, regardless of whether or not it is permissible on paternalistic grounds. It is clear, however, that not all harm to others is sufficient to justify intervention (see Feinberg, 1984), and doubtless it would be a very rare case of piano practise in which the overriding moral consideration is preventing harm to others.

The justification of compulsory education is related to this kind of case, but differs from it in two important ways. Because the benefit to children of a certain amount of education far outweighs in our society the benefit of practising piano lessons, there is a greater chance that B will subsequently approve_e of prior paternalistic interference that enabled him to become literate and at least somewhat informed. Also, there are particularly strong non-paternalistic grounds on which compulsory education can be justified, again on the grounds of preventing harm to others. The significant others in this case are the members of B's society who, it is often said, will be harmed, at least financially and perhaps in other ways, if B grows up illiterate and unskilled. (Education is discussed further on pp. 204-208 below.)

(9) Paternalistic Indoctrination

T, who is seventeen years old, is kidnapped by members of a cultish religious sect who condition and indoctrinate him into a strange way of life centering on the worship of various species of non-human animals and plants. Within a few months T's parents re-kidnap him, "de-programme" him, and T returns to a normal way of life.

Here we have two cases of paternalistic interference. The first instance of indoctrination leading to T's religious conversion is, in the absence of extenuating circumstances, morally uncontroversial. The indoctrination is unjustified because it inhibits not only the exercise but also the development of T's autonomy. T's deprogramming, however, may use exactly the same methods in bringing T back to his normal state as the religious sect used on T in the first instance. Some people may feel vaguely uneasy about both approving of such methods for re-indoctrination into a normal way of life, and disapproving of them when used for religious ends, as if there were some inconsistency here. But there need not be any inconsistency. Some degree of minimally autonomous approval_e can be reasonably expected on T's part subsequent to the deprogramming, and therein lies the justification for it. The deprogramming leads, hopefully, to greater reflective ability on T's part than he was previously able to exercise. T may be blatantly indoctrinated by his deprogrammers into the view that the religious group which kidnapped him are not the enlightened persons he thought they were, but instead highly manipulative

cult figures, so that initially T holds this view not on the evidence (for at first he is in no mental state to assess the evidence), but because the deprogrammers have drilled it into him. But presumably he later comes to hold it for reasons which are causally efficacious in his maintaining it, reasons which pertain to his assessment of the evidence. If so, his autonomy is restored. By contrast, the beliefs with which T was religiously indoctrinated presumably never would have come to be held on evidence, and would always have prevented T from achieving greater reflective ability, and hence, autonomy.

It does not follow from this that any indoctrination of children is permissible provided the children later adopt the indoctrinated beliefs for reasons, and those reasons are causally efficacious in the maintenance of those beliefs. Minimally autonomous approval_e of the indoctrination is also required. Only in very rare cases do people autonomously approve of their own prior indoctrination.

(10) Paternalistic Deprivation of Opportunity

The Amish are descended from eighteenth century immigrants of strong Protestant conviction who settled in the U.S. in order to organize self-sufficient farming communities along religious principles, free of interference from unsympathetic outsiders. The aim of Amish education is to prepare the young for a life of industry and piety by transmitting to them the unchanged farming and household methods of their ancestors and a thorough distrust of modern techniques and styles that can only make life more complicated, soften character, and corrupt with "worldliness". Accordingly, the Amish have always tried to insulate their communities from external influences,

including the influence of State-operated schools. Their own schools teach only enough reading to make a lifetime of bible study possible, only enough arithmetic to permit the keeping of budget books and records of simple commercial transactions. Four or five years of this, plus exercises in sociality, devotional instruction, inculcation of traditional virtues, and on-the-job training in simple crafts of field, shop, or kitchen are all that is required, in a formal way, to prepare for the traditional way of life to which Amish parents are bound by the most solemn commitments.

In the case of Wisconsin V. Yoder (et al. 406 U.S. 205, 1972), the Amish won a resounding victory in the Supreme Court of the United States. The Amish litigants in that case had been convicted of violating Wisconsin's compulsory school attendance law (which requires attendance until the age of sixteen) by refusing to send their children to public or accredited private school after they had graduated from the eighth grade. The U.S. Supreme Court upheld the Wisconsin Supreme Court's ruling that application of the compulsory school-attendance law to the Amish violated their rights under the Free Exercise of Religion Clause of the First Amendment.

(Adapted from Feinberg, 1980)¹⁴

The paternalistic interference on the part of the Amish parents in this case was towards those children whose right to attend school was violated by their parents' insistence that they attend no further than the eighth grade. This group of children includes both those who wanted to continue their schooling beyond the eighth grade, and those who were in favour of their parents' decision. The former suffered an infringement of their rights to non-interference₁ and ₃, and the latter suffered an infringement of their right to non-interference₃ only.

The majority opinion in this case, written by Chief Justice Berger, favoured the Amish parents not on the grounds that this was a case of justified paternalism, but on the grounds that

if the State is empowered, as parens patriae, to "save" a child from himself or his Amish parents by requiring an additional two years of compulsory formal high school education, The State will in large measure influence if not determine, the religious future of the child ... This case involves the fundamental interest of parents, as contrasted with that of the State, to guide the religious future and education of their children.

(Feinberg, op.cit.)

Feinberg points out, rightly, that there is a mistaken assumption behind this line of thought, the assumption that any influence by the State in this context constitutes undue influence on the religious future of the child. The mistaken assumption is that neutrality is impossible. Although Feinberg does not support Burger's reasoning, he does, however, support the Court's finding in favour of the Amish parents on the grounds that

The difference between a mere eight years of elementary education and a mere ten years of mostly elementary education seems so trivial in the technologically complex modern world, that it is hard to maintain that a child who has only the former is barred from many possible careers while the child who has only the latter is not ... the difference between eight and ten years is minor in terms of the children's interests, but possibly crucial for the very survival of the Amish sect.

(Op.cit.)

Feinberg's point is that the interests of the child in having an open future are paramount, but the extra two years of schooling at stake in Yoder would not significantly contribute to a more open future for Amish children, and it would have a corrosive effect on the religious community. Feinberg is right, no doubt, that the difference between eight and ten years of schooling would make little difference to the future career opportunities of Amish children. He is probably also right that the extra two years of schooling in the state high school would be effective in loosening the commitment of Amish children to the traditions of their religious community. That is because the

secular influences on the children had been minimal during the first eight years since they attended a 'nearby rural schoolhouse', with an overwhelming proportion of students of the Amish faith, none of whom played rock records, watched television, or the like ...
(Op.cit.)

But the secular influence of non-Amish children on Amish children could be expected to be substantial at the secondary school.

Feinberg weighs up what he takes to be the significant disvalue to the Amish community of sending their children to state high schools against the insignificant increase in career opportunity for Amish children, and decides in favour of the Amish community. But Feinberg is a bit hasty in concluding that because the future career interests of Amish children will not be enhanced by two extra years of schooling, their deprivation

of opportunity in this case is not morally significant. He admits that the religious bonds of the Amish community could be significantly diminished if Amish children attended state high schools. That in itself would strongly affect the future of the children concerned, quite apart from questions of career. Depriving them of the opportunity to witness other ways of life may not be something they could subsequently approve, freely, particularly if they remained unaware of possible alternative ways of life. At the same time, perhaps the contribution to greater autonomy in terms of greater awareness of different ways of life the Amish children might acquire if they attended state high schools, would be offset by a diminution of autonomy in terms of diminished ability to be faithful to their chosen Amish way of life, i.e. diminished self-control. In any case it remains an open question whether, if the finding had been paternalistic, the Amish children would subsequently freely approve_e of this deprivation of opportunity.

Again, there is an alternative non-paternalistic line of defence for Yoder. If the harm caused, not to the Amish way of life (for the Amish way of life is not a being and may not in itself have moral significance) but to the Amish parents by their children's attending a secular high school, were weighty enough, then the deprivation of opportunity for their children might be defensible on the grounds of preventing harm to the parents. It cannot be assumed, however, pace Feinberg, that the loosening of the religious bonds of the

Amish community is necessarily a harm. If harms are the dissatisfaction of informed desires, then although it is clear that the desires of Amish parents for themselves and their children are very strong, it is not so clear that they are informed. Many Amish parents may themselves have been permanently indoctrinated. In this kind of case we would need to satisfy ourselves that their desires were in fact informed before allowing them to carry this kind of moral weight.

(11) Temporary Incompetence

W is rushed to hospital after accidentally ingesting a lethal dose of poison from a mismarked bottle. The poison makes him delirious, and he refuses to drink the harmless antidote the doctors give him, claiming they are trying to kill him. They subdue him, and force him to take the antidote intravenously.

In this comparatively straightforward sort of case it is clear that the patient is not competent to make his own decision about whether to take the antidote. The patient's incompetence is obvious evidence that intervention against his expressed will is justified, for he will no doubt later approve_e of this intervention and see himself as having been temporarily incompetent at the time of treatment. If his refusal of the antidote were based on deeply held religious beliefs, however, then it would be a very different sort of case, in spite of the fact that his incompetence at the time of intervention remains the same. Temporary incompetence is often

good evidence that beneficial intervention will be subsequently consented to, but it is not conclusive evidence. The overall settled preferences of the individual must be taken into account also.

Interestingly, some Jehovah's Witnesses appear to have a loophole in their beliefs which may permit physicians to give a blood transfusion to a non-consenting Jehovah's Witness. These are the patients who want to live, yet cannot consent to a transfusion, and who believe that a court-ordered transfusion would not affect them adversely either morally or religiously. Childress suggests that "In a few cases it may be a matter of both beneficence and respect for persons to help such patients through the loophole in their beliefs" (1982, p.169). Whether there is a loophole depends on whether it is against the Jehovah's Witness's will that he receive a blood transfusion, or whether it is only against his will that he himself give permission for a blood transfusion. Probably most Jehovah's Witnesses would claim the former position as theirs, but some claim the latter. If W were one of the latter and needed a blood transfusion, it might be moral cowardice on his part, but his position is compatible with his wanting the doctor to take matters into his own hands and administer a blood transfusion without W's permission. If the doctor does so, he does not violate W's right to non-interference because W does not object to the doctor's giving him the transfusion. He objects only to giving permission for it. In such a case W's subsequent approval_e of the doctor's action

is to be expected in spite of the fact that explicit consent at the time of the transfusion and afterwards will be withheld.

Some commentators have argued that even in cases when there is no loophole of this kind in the beliefs of a Jehovah's Witness, he ought to be given a blood transfusion if that is necessary to save his life on the grounds that this would be in keeping with the subject's "objective" wishes (Glass, 1983, p.475). Objective wishes are those an agent would have if fully informed. A Jehovah's Witness would in fact want a blood transfusion, Glass says, if he knew he was mistaken about what the Bible requires. That is because it is true of all of us, he adds, that we prefer to have our informed desires satisfied even when they conflict with present interests (p.474). This kind of "real will" argument purports to be autonomy-respecting, but is in fact autonomy-defeating. It ignores the fact that a Jehovah's Witness, like everyone else, regards freedom of worship and all that that entails as very much in his interests.

(12) Selling Oneself into Slavery

It is sometimes thought that the non-enforcement of slavery contracts is a hard case for someone who does not wish to endorse strong paternalism. Mill says, of someone who would voluntarily sell himself into slavery:

By selling himself for a slave, he abdicates his liberty; he foregoes any future use of it beyond that single act. He therefore defeats, in his own case, the very purpose which is the justification of allowing him to dispose of himself ... The principle of freedom cannot require that he should be free not to be free.

(On Liberty p.125)

Hence Mill is opposed to a law which would enforce slavery contracts. But his reason for opposing it, that the principle of freedom cannot require the freedom not to be free is not a good one. For entering into any short or long term contractual agreement limits one's freedom to some extent, and Mill is not opposed to other kinds of contractual agreements.

A better argument for the non-enforcement of slavery contracts which is sometimes put forward, is that it would be very difficult to guard against exploitation if slavery contracts could be legally enforced (Feinberg, 1971, pp.118-120). However, this is an argument based on expediency. It does not suggest that there is any reason in principle for the non-enforcement of slavery contracts. Hodson (1981) argues, alternatively, that there is no need to justify the non-enforcement of slavery contracts on the grounds that there is no *prima facie* objection to the non-enforcement of slavery contracts. The government's failure to enforce slavery contracts does not encroach on anyone's liberty, suggests Hodson. That is why there is no *prima facie* objection to it. There are many services the government does not provide, he adds, and this is one of them.

However, even if the non-enforcement of slavery contracts is not an encroachment of anyone's liberty rights, as Hodson says, it may nevertheless be an infringement of some other right. Given that the protection of contracts is an important legal function, it could be argued that any contract voluntarily entered into which affects only the participants ought *prima*

facie to be upheld in the courts. So there might be at least a prima facie right on the part of either slave or contractor to have a slavery contract upheld. But as Feinberg points out, there are good grounds for the presumption that anyone entering into such a contract is acting involuntarily; and voluntariness would be too difficult and expensive to prove in court. So arguments from expedience are clearly significant.

The non-enforcement of slavery contracts is not the only example of apparent paternalism in the law of contracts. Other examples include the invalidity of contracts made by minors, the invalidity of arguments purporting to waive the promisor's right to obtain a divorce, and the invalidity of provisions in some residential leases that purport to waive the tenant's right to withhold payment of rent if the residence fails to meet certain standards of habitability. In the latter case the law protects the tenant by refusing to enforce his waiver. Kronman (1983) has argued that there is no single principle that best explains every paternalistic restriction in the law of contracts. Some restrictions are best explained, he suggests, by considerations of economic efficiency and distributive fairness, others by the idea of personal integrity (e.g. the invalidity of slavery contracts), and others by the notion of sound judgement (e.g. the invalidity of contracts between minors). However, there is no reason to believe these diverse principles are independent of the unifying principles we have suggested for the justification of paternalistic

intervention. Indeed, if they are acceptable explanations of paternalism in the law of contracts, then according to our theory they must be subsumable under the principles of justification we have been exploring (on the assumption, that is, that they constitute a *prima facie* violation of the recipient's autonomy).

Discussion of slavery contracts is also an area where the protection of future selves seems relevant. What moral right has a fifteen-year-old boy, it could be asked, who voluntarily enters into a slavery contract, to bind the sixty-five year old he will eventually become? According to Feinberg's (1971) very high standards of voluntariness, it is doubtful that such standards could be met by anyone entering into a slavery contract to begin with.¹⁵ But if we allow that a fifteen-year-old boy could enter into a life-long slavery contract voluntarily, our objection to such a contract is best accounted for by our concern for the autonomy and well-being of the future self this adolescent will become. Thus preventing harm to others, namely future selves, may also be good grounds, at least on the reductionist view of personal identity, for the non-enforcement of slavery contracts.

(13) Paternalistic Lying

Q, on the advice of Dr. R, enters the hospital for routine tests after suffering for several months from abdominal pain, loss of appetite, and a number of other symptoms. During exploratory surgery, Dr. R discovers that Q has cancer of the liver, and he estimates that Q has about three months to live. Dr. R informs Q's wife, and together they decide not to tell

Q of the gravity of his condition. Instead, Q is informed that his symptoms are the result of duodenal abnormalities which should clear up after a few weeks of rest and medication. Q is delighted and sets about writing a book, after having felt disinclined to work for several months.

This case is different from the other cases of paternalism we have considered so far because it does not involve the recipient's right to non-interference. If the paternalistic lie violates one of Q's rights, it violates his prima facie right not to be deceived. So far we have ignored cases of paternalism which constitute violations of some right other than the right to non-interference because, unlike the right to non-interference, it is not clear which other rights can be alienated by the subject himself. But if the right not to be deceived can be alienated by the approval_e of the possessor of the right, then for our purposes the right not to be deceived is very similar to the right to non-interference.

It could be argued that there is an intimate connection in this case between the right not to be deceived and the right to non-interference₃ (non-interference with the recipient's opportunities). Not knowing the truth about his condition deprives Q of many important opportunities he might otherwise have - opportunities both to get his affairs in order and to share the experience of dying with his spouse and family. In addition, if Q has a right to know the truth, depriving him of it interferes with his belief states. So there is also a sense in which this deprivation could count as a violation of

the right to non-interference .

2

In any case, the points to be considered are whether the deception benefits Q, and whether he himself somehow fundamentally prefers to remain ignorant. Since benefit is the satisfaction of informed desire, it may seem that this kind of deception has little chance of benefitting Q, for in deceiving him it prevents his informed desires (the ones he would have if he knew his condition) from becoming actual desires. However, it is possible to attribute to Q an informed desire not to be informed of his impending death. If such is the case, Q can be said to alienate his right not to be deceived in such a situation. If Q would prefer to know the unhappy truth in spite of the depression he would subsequently suffer, then he does not alienate his right not to be deceived, and the paternalistic lie is unjustified. Not everyone is the sort of person who prefers to know the truth in such circumstances. Only someone who knows Q very well would be competent to judge whether he prefers the benefits of ignorance to the pain of knowing the truth about himself. Since so much depends on the kind of person Q is, it is impossible to generalize about the justifiability of paternalistic lies in such cases. Some persons would readily alienate their right not to be deceived in such circumstances, though perhaps most would not. The point of this example is to show that in cases where paternalism is a violation of some right other than the right to non-interference, the conditions of justification might remain the same, though in some cases they might be much more

difficult to apply.

A similar kind of case to paternalistic lying is the paternalistic breach of confidentiality. If, for example, a secondary school counsellor reneges for paternalistic reasons on an assurance to students of confidentiality, when confronted with cases of, say, drug and alcohol addiction, she is not necessarily violating their right to non-interference, though she is obviously violating their right to confidentiality. There are frequent cases of the justified breach of confidentiality on the part of priests, psychiatrists, lawyers, and other professionals for the purpose of preventing serious harm to others (Bok, 1984, Chpt.9), but it is doubtful that there are many cases of justified paternalistic breach of confidentiality because those who confide would only in very exceptional circumstances subsequently approve of such a breach. For that reason, professionals should be careful not to promise confidentiality in cases where it might conflict with some other duty.

2. Legal Paternalism

The difficulties of ensuring that paternalism is autonomy-respecting are greatly increased in the legal realm. Special problems arise where large numbers of people are concerned. Compulsory treatment of heroin addicts, for

example, may benefit some, but not all, or even most heroin addicts. But even if compulsory treatment were always a positive benefit, not all addicts would alienate their right to non-interference with respect to it. The situation is complicated by the fact that others, besides addicts, would also be affected by a law requiring compulsory treatment. Non-addicts would be indirectly affected by having to finance, through taxation, a scheme of compulsory treatment for addicts, and also by the whole social and political atmosphere the legislation may help to bring about. Of course addicts themselves would be affected not only directly by subjection to compulsory treatment, but also in the ways affecting non-addicts.

It is clear that inter-personal paternalism is sometimes justified qua paternalism, while unjustified from an all-things-considered point of view. For example, it might sometimes be justified from a paternalistic point of view to impose music lessons, say, on an unwilling child, but unjustified from an all-things-considered point of view if the cost of those lessons is so high it deprives other members of the child's family of basic needs. However, this distinction between paternalism justified qua paternalism and paternalism justified from an all-things-considered point of view is inapplicable in the legal realm. There are two reasons for this. First, the cost of implementing and enforcing paternalistic policies nearly always substantially affects persons other than those meant to benefit from them. So the question of other people's

interests is nearly always relevant. Second, even if the interests of persons indirectly affected by a paternalistic policy could be ignored, it is highly unlikely that all persons would benefit and approval_e will be unanimous among those directly affected. If we suppose thirty per cent, say, of all heroin addicts either would not benefit from compulsory treatment or fail to alienate their right to non-interference with respect to it, the question is whether we can justify the coercion of this thirty per cent (or forty or sixty as the case may be) on non-paternalistic grounds for the sake of those who would in fact benefit from the policy in question and also alienate their right to non-interference. So again, the question of preventing harm to others impinges from the start. These considerations suggest there will be few cases of legal paternalism, especially criminal paternalism, justified from an all-things-considered point of view. There is, however, no objection in principle to legal paternalism as such, when it provides autonomy-respecting benefit.

Often what appears to be a paternalistic policy does not actually constitute objectionable interference. If competent persons consent to having their liberty restricted in a certain way for their own good, there may be no violation of rights and hence no paternalistic interference. Laws against duelling, for example, as well as seatbelt legislation probably fall into this category. Most people want them, and those who object to having their liberty restricted in these ways can be coerced

(not on paternalistic grounds for paternalism may not be autonomy-respecting in such cases) but on the grounds of preventing harm to others.

Criminal paternalism makes certain self-regarding harmful acts punishable by the imposition of fines, jail sentences, and criminal records. Suicide as a criminal offence is one example of criminal paternalism. Other commonly cited examples include laws regulating the use of certain drugs which may have harmful consequences for the user but do not lead to anti-social conduct, laws forbidding various kinds of gambling, laws regulating certain kinds of sexual conduct, laws requiring motorcyclists to wear safety helmets, and so forth. Sometimes the intent behind such laws is only partly paternalistic. The intent behind motorcycle helmet and seatbelt legislation, for example, may be in part for the protection of accident victims, and in part to curb the high cost of medical treatment for them.¹⁶

Bayles (1974) argues that paternalism is not a morally acceptable reason for criminal legislation. In trying to justify criminal legislation, he holds, one must not only show that a person's action may be restricted and that criminal sanctions are an effective means to secure their restriction, it must also be shown that criminal paternalism is less harmful than any other means. To this it could be added that it must also be shown that it is permissible to restrict non-beneficiaries who lose liberty or otherwise suffer a violation of rights by the legislation in question. Concerning the first of these criteria -

showing that a person's actions may be restricted - Bayles takes the standard liberal position that weak paternalism is morally justified but strong paternalism is not (i.e. that interference is justified only where the will of the beneficiary is incapacitated in some way.) Many people would agree with Bayles that since failing to wear a seatbelt is a voluntary action, there is no acceptable paternalistic reason for restricting liberty in this regard. He points out that people may indeed consider the risk worth taking. However, the voluntariness of a driver in failing to buckle up is relevant only if it constitutes good evidence that he himself does not, and will not, approve of having his liberty restricted in this way. It is possible, however, even probable, that most car drivers and passengers approve of having their liberty restricted in this way. Many people feel psychologically more secure wearing a seatbelt and knowing it is compulsory to do so. Many are aware of benefitting from buckling up even though they may never be involved in a car accident. This point tells against the once frequently heard complaint that it is unfair to restrict the liberty of everyone for the sake of preventing harm to accident victims whose own fault it is for not buckling up. In addition, it suggests that most people want seatbelt legislation, and hence it is not properly regarded as paternalistic interference.

Besides the argument from weak paternalism, Bayles offers a second reason why actions, voluntary or not, should not be restricted by criminal paternalism. "There is a sort of

inconsistency," he says, "between the motive of avoiding injury to actors and intentionally inflicting injury on them" (1974, p.184). But why should it be thought there is a sort of inconsistency between the motive of avoiding injury to actors and the motive of intentionally inflicting injury on them. Inflicting injury on someone may sometimes be a means of avoiding greater injury to him, and there is nothing inconsistent about that. Inflicting injury on someone through legal sanctions for his own good, may be largely ineffective in enhancing his well-being, but it is no threat to consistency.

Bayles also claims that criminal paternalism is not an effective means to the restriction of self-harming conduct. Educative means, he suggests, would be generally more effective and less harmful than criminal paternalism. This double empirical claim is difficult to apply across the board, however, as Bayles wants to do, to all cases of criminal paternalism. It is not the case that the "educative means" of getting people to wear seatbelts which was employed before seatbelt legislation became effective, was nearly as effective as the subsequent threat of financial penalty. Still, Bayles general point is well taken that to justify criminal paternalism one must not only show that a person's action may be restricted and that criminal sanctions are an effective means to secure their restriction, it must also be shown that criminal paternalism is less harmful than any other (e.g. educative) means.

Some commentators have gone to very great lengths to

defend motor-cycle helmet and seatbelt legislation on both paternalistic and non-paternalistic grounds. It has been suggested, for example, that failing to buckle up is not fully voluntary (Lee, 1981). If someone is incapable of taking very small probabilities into account, as most people are, this may be regarded as a kind of ignorance, and to act on ignorance, so Lee's argument goes, is to act involuntarily. But this attempt to justify seatbelt legislation depends on an implausibly high standard of voluntariness. It has also been suggested that motor-cycle helmet legislation can be justified on the grounds of preventing harm to others, the others being one's future selves (Regan, 1974). In putting this view forward, Regan relies on an unorthodox view of personal identity, in which someone may be conveniently regarded as a different person for some purposes but not for others. A motorcyclist who voluntarily refuses to wear a helmet becomes a different person in one respect after an accident, says Regan, if he is no longer the sort of person who would ignore his future well-being for the sake of small increments of present utility, although he may not be a different person for other purposes. But we need not rely on such arguments to defend this kind of legislation. With seatbelts, as we have said, most people appear to forfeit their liberty willingly in this way. If they did not, then enforcing the use of seatbelts by the threat of legal sanctions would not be autonomy-respecting. Subsequent approval_e in cases of this kind is unlikely unless

there is willing forfeiture of liberty to begin with. If people were opposed to it, seatbelt legislation might still be defended on other grounds, however, especially if the saving in medical costs were in fact substantial enough to warrant this kind of coercion. But then of course the justification would not be paternalistic.

Ten (1980) has pointed to an apparent inconsistency between seatbelt legislation on the one hand, and on the other the absence of legislation restricting mountain climbers. It might be thought that the benefits to the mountain climber of the sense of exhilaration and achievement which motivates his activity, provides far greater benefit to him than any benefit that might accrue to drivers who prefer to take risks by not buckling up. In answer to this Ten argues that we have no way of distinguishing between benefit to the mountain climber of the risks he takes and benefit to the driver who does not buckle up (1980, pp. 116-117). Ten is correct in suggesting we have no reliable way of distinguishing between benefit to the mountain climber of the risks he takes and benefit to the driver who wants not to buckle up, and clearly we should not merely assume benefit is greater for the former than for the latter. But from this Ten concludes we are inconsistent in restricting car drivers through seatbelt legislation while not restricting mountain climbers. However, the apparent inconsistency disappears if we regard seatbelt legislation as a case of the willing forfeiture, by most people, of their liberty in this regard, whereas it is

very doubtful that mountain climbers would willingly subject themselves to similar sorts of government regulations.

Perhaps a stronger parallel holds between mountain climbers and motor-cyclists, than between mountain climbers and car drivers. If it were thought (plausibly or implausibly) that unlike car drivers, motor-cyclists prefer taking greater risks and do not generally approve of legislation requiring them to wear safety helmets, it could be asked what grounds there might then be for defending this kind of legislation. Probably most victims of motor-cycle accidents subsequently approve_e of this restriction, as do many motor-cyclists who have not had accidents, but it is doubtful that coercion of the rest can be approved on the grounds of preventing harm to only this small number of motor-cyclists. This suggests there may be some inconsistency between restricting motor-cyclists on the one hand, but not mountain climbers on the other. However, many others besides the accident victims themselves suffer when motor-cyclists fail to wear helmets. If we take full account of the harm caused to others in this way - not only the high cost of medical treatment, but also the offended sensitivities of people from the frequent occurrence and sight of gory accidents - there are probably better non-paternalistic grounds to justify interfering with the liberty of motor-cyclists, and requiring them to wear safety helmets, than there are with mountain climbers.

A similar non-paternalistic case could perhaps be made

for the compulsory treatment of heroin addicts, depending on the effectiveness of the treatment, given that heroin addiction often leads to stealing and other forms of anti-social conduct in order to support the addiction. Compulsory treatment could be paternalistic in intent, but without these other non-paternalistic grounds to recommend it, it is doubtful that the autonomy-respecting benefit that some addicts might obtain from it would be weighty enough in itself to justify coercing the rest.

Similar points can be made in defense of certain kinds of so-called "impure" paternalism. Legislation which prohibits the manufacture and sale of harmful drugs is an example of impure paternalism in which the protected party is different from the party interfered with. As Dworkin (1971) points out, cases such as these are different from ordinary non-paternalistic cases of preventing harm to others (e.g. legislation which restricts a manufacturer from polluting the air) because the free choices of the protected party are ignored. That is what makes them paternalistic. Bayles (1974) disputes this on the grounds that paternalism requires that A's action cause injury to himself. Thus, legislation prohibiting the manufacture and sale of harmful drugs is not, according to Bayles, paternalistic. However, such legislation need not fall wholly into one category or the other. In so far as it curbs the liberty of drug manufacturers, it must be defended on the grounds of preventing harm to others; and in so far as it curtails the liberty of

potential drug users, paternalistic considerations are relevant, though not necessarily conclusive.

One example (hypothetical) of unjustified impure paternalism might be legislation against the sale and manufacture of cigarettes. Because it is doubtful that would-be smokers are likely to alienate their right to have their free choices restricted in this way, the possibility of autonomy-respecting paternalism in this kind of case is eliminated. And since there are only very weak non-paternalistic grounds for banning smoking in private, we must conclude that the manufacture and sale of cigarettes should continue. (The advertising of cigarettes, however, is another matter, and there is good reason to ban it on the grounds of preventing exploitation.) If the sale and manufacture of very dangerous drugs such as heroin were in question, however, the conclusion would be different. Because many people feel threatened and may be harmed even by witnessing the recent increase in our society of heroin addiction, there are strong non-paternalistic grounds for making the use of heroin illegal. These same grounds support banning the sale and manufacture of it too. Besides, most people probably willingly implicitly consent to the existence of laws which interfere with their freedom to purchase and take heroin or other dangerous substances, simply because of the devastating physical and psychological effects such drugs can have. This kind of legislation can probably also be defended, therefore, in the same way as laws against duelling or seatbelt legislation - on

the grounds that it reflects the actual preferences of most of the people to whom the laws apply.

There are, of course, very large questions lurking in the background here, concerning whether and for what reasons a society may prohibit individuals or groups of individuals from engaging in practices which the majority find threatening, offensive, immoral, disgusting and so forth. In relation to issues about "private" morality and certain so-called victimless crimes, Devlin's (1959) famous response to the Wolfenden Report on homosexual offences and prostitution, Hart's (1959) response to Devlin, and more recently Dworkin's (1981a) response to the Williams Report on obscenity and film consorship, address many of these questions which, for present purposes, must be set aside. However, it is worth mentioning - if only because laws restricting pornography and the like are sometimes thought to be examples of paternalistic legal moralism - what the harm could be that legal moralists might claim is prevented by, for example, the restriction of pornography.

Here legal moralism is understood to be the view that immorality should sometimes be prohibited by law with the understanding that this is an independent principle from that of preventing harm to others. Harm to others usually means any of the following: 1) direct physical damage to people, their property, or financial interests, 2) mental distress, or 3) damage to the general social and cultural environment (Dworkin 1981a, p.178). The harm at issue here in considering paternalistic legal moralism, however, is the moral harm that is purportedly prevented from

occurring to those who might otherwise read or enjoy pornographic material. This idea of harm has nothing to do with well-being as defined in Chapter Three, or its opposite - the dissatisfaction of informed, fundamental desires - because legal moralism may hold that an agent is morally harmed by acquiring certain informed, fundamental desires rather than others. Legal moralism, we might say, presupposes a full conception of the bad.

Now, we said in Chapter Three that as long as paternalistic interference increases the recipient's informed, fundamental desire satisfaction (and assuming the interference is autonomy-respecting), a paternalist might permissibly choose to promote the satisfaction of some desires at the expense of others because of his belief that certain desires are more in keeping with human flourishing than others. Similarly, it might be thought that the state, as paternalist, could permissibly discourage or prevent the development of certain tastes or desires in citizens if this could possibly be done in an autonomy-respecting way which also enhanced the well-being of citizens. In principle there may be nothing objectionable about this kind of legal moralism qua autonomy-respecting paternalism. Indeed this is just how responsible parents act towards their children a good part of the time. However, few of us would accept a state that acted towards its citizens as a parent towards children. Most of us believe the state should remain as neutral as possible among competing conceptions of the good, or as the case may be, among competing conceptions of the bad, although we do not think it appropriate

for parents to be neutral in this way with their children. For this reason, legal moralism as such is unacceptable.

Besides criminal paternalism, it is worth mentioning that there is in our society a considerable amount of non-criminal paternalistic legislation. Non-criminal paternalistic legislation usually relates to minors. Laws prohibiting minors from entering into hire-purchase agreements, and laws prohibiting children from accepting certain kinds of employment, protect them against possible exploitation by potential employers. However, because adults have the right to have hire-purchase agreements upheld by the courts, and to seek (with certain obvious exceptions) whatever kind of employment they wish, depriving minors of these rights requires justification. For as we have said, there is no qualitatively morally relevant difference between children and adults which recommends different treatment for each group. But because children are more easily exploitable than adults, there is little doubt that such legislation is of benefit to them. Furthermore, since few adults disapprove_e of having had their freedom restricted in these ways as children, most people readily alienate their right to non-interference with regard to such laws.

3. Education

Finally, it is worth discussing separately one important area of legal paternalism - compulsory education - which generally

affects everyone and is normally believed (correctly we may assume) to be justified.

Generally speaking, compulsory education in our society benefits recipients to varying extents, and also to varying extents subsequently meets with their approval_e. Many so-called progressive educators have often endeavoured to make compulsory education autonomy-respecting by adopting curricula and methods of teaching based on children's present interests and desires about what they should learn. But as we have seen, respect for the autonomy of children need not be based on permitting children's present interests causally to determine present activity. Regardless of the differences between progressive and less progressive educators, however, the development of intellectual autonomy is considered by both to be an important educational aim, whether they permit pupils in classrooms to exercise very much overt freedom of choice or not. The following remarks are made with the development of autonomy in view as an important educational aim.

The pursuit of knowledge and understanding in the many fields of human achievement is normally highly valued for its own sake among persons with first hand experience of exploring one or more disciplines in depth, and of coming to see and feel differently about the world because of this. But for many persons the pursuit of knowledge is valued largely for instrumental purposes - its contribution towards vocational training, advanced technology, and so forth. Where the acquisition of knowledge is regarded as

largely instrumental (as it frequently is even among educators, perhaps more so in North America than in Britain), the demand is frequently heard that what is learned should be relevant in some specifiable way to the actual political, social and personal problems students will encounter in their lives. One can sympathize with the feeling behind this demand that knowledge ought to make some significant difference to an agent's life. But of course the differences knowledge makes are not always perceptible or specifiable in advance as directly relevant to finding solutions to present problems. However, even instrumentalists have a good reason in spite of this for valuing the pursuit of knowledge independently of its instrumental relationship to other purposes, if they value autonomy as an intrinsic good, because as we have seen knowledge is required for autonomy.

Furthermore, if the development of autonomy comes about through learning and not simple maturation, and autonomy is taken as an important educational aim, then educational institutions which promote the pursuit of things like computer technology, vocational training, literacy but not literature, games, crafts, and so-called life skills, at the expense of a liberal education, short-change learners in an important way, though not necessarily in a way that is experienced by the learners themselves as being short-changed. The notion of a liberal education, as it is used here, roughly refers to what Warnock (1978) has called the education of the imagination. She says:

What I am urging is that the goal of education should be to understand and interpret a whole range of phenomena,

natural and man-made, objects of mathematical, mechanical, aesthetic interest alike, so that within the particular field of study, whatever it is, the feeling of there being more, more to learn, more to think about, more to try to do, is the predominant feeling.

(1978, p.57)

This feeling is familiar to someone deeply immersed in one or more disciplines, but it is not available in full measure to instrumentalists.

Finally, many so-called progressive educators have taken the view that the way to develop a child's autonomy is to give him free rein in exploring what he wants, but our analysis suggests that the development of autonomy is better served by helping him improve his ability to think competently about many different subjects, including his own values. This is achieved by helping him to examine critically what others think, and have thought in the past, as a foundation for, and an impetus to, his forming views of his own. It makes good sense, of course, to allow children to develop a sense of their own abilities by allowing them to make choices in classrooms and be responsible for them; but that is only one small part of the development of autonomy, a part which progressive educators have often stressed at the expense of the development of deeper reflective abilities.

Besides, with regard to self control and moral autonomy, if Kohlberg (1969) is right that children pass through invariable stages in their perception of rules before they reach the autonomous stage, and that successful passage from one stage to the next depends on the kind of cognitive stimulation the child receives, then it is futile to expect that a very young child

who looks at rules largely in terms of reward and punishment, can learn to feel responsible by giving him full free rein. It may be necessary to experience something of firm guidelines first, in order to acquire some understanding of what it is to obey a rule, before the agent attempts the more difficult task of imposing his own rules on himself.

This is not to suggest that old-fashioned teaching methods with emphasis on rote learning and punishment deserve any significant place in the development of reflective ability. No doubt they are highly counter-productive. But it is important that educational institutions, at the lower as well as higher levels, promote, as much as possible, reflective understanding of and in the major fields of human knowledge.

4. Conclusion

We have defined paternalism as any prima facie violation of a recipient's right to non-interference for the purpose of protecting or promoting the recipient's well-being. As a starting point for the thesis justified paternalism was taken both to respect the autonomy of the recipient and to enhance his well-being. A theory of individual autonomy was developed in Chapter One emphasizing the importance of self-construction and self-control. Rational reflection was shown to be central to the idea of self-construction. In Chapter Two the right to non-interference was extended to all beings with desires. It

was argued that a *prima facie* violation of a recipient's right to non-interference is a non-actual violation if the recipient alienates his right through subsequent approval_e. A conception of personal well-being was adopted in Chapter Three in terms of the satisfaction of informed, fundamental desires. It was seen that any substantial enhancement of a recipient's well-being will either meet with that recipient's approval_e, or seriously fail to be autonomy-respecting. This led to the consideration of some important questions about perfectionism and well-being. In this, the fourth chapter, certain general implications of the theory of autonomy-respecting paternalism have been discussed pertaining to judgements of incompetence and the role of consent in the justification of paternalism. Also, the morally relevant factors in several specific cases of paternalism have been considered.

The main task of the thesis has been to describe the conditions or states of affairs which must obtain in order to satisfy the requirements of justified paternalism, namely that it respect the autonomy of the recipient and enhance his well-being. When paternalism satisfies these conditions, it is, we have said, objectively justified. The findings of the study provide a framework from which the theoretical relevance of consent, and judgements about a recipient's competence or incompetence in the context of justifying paternalism can be readily seen. Perhaps the most surprising finding is the conclusion that paternalistic interference with a competent,

unwilling subject acting voluntarily, can be autonomy-respecting if the subject subsequently alienates his prior prima facie right with respect to the interference. Such cases are rare, and rarer still is the ability accurately to predict which ones they are. However, the view that such cases are possible has two satisfying results, one practical and one theoretical.

The theoretical result is that we no longer need the double standard of the traditional approach to the justification of paternalism. The traditional approach emphasizes respect for the autonomy of adults only. Our theory of autonomy-respecting paternalism applies to everyone, including children. The practical result is that parents and educators who tend to feel uneasy about paternalistically overriding children's voluntary choices, can be assured that they are not necessarily abandoning respect for the autonomy of children in such cases.

There remain several areas of the thesis which, on another occasion, could be usefully explored further. They include the following: 1) explicating the notion of minimally autonomous approval_e introduced in Chapter Two, 2) discussing in much greater detail which specific rights to non-interference we have, and 3) discussing in detail the specific institutional guidelines which in different contexts (e.g. doctor/patient relationships) best achieve the objective conditions of autonomy-respecting paternalism. In addition to these questions, the larger issue of paternalistic duties needs much attention.

So far most writers on the justification of paternalism have devoted their attention to the work of outlining various moral and institutional guidelines paternalists ought to follow, at what we have called on page 18 justification levels two and three. Often, however, guidelines are given without much theoretical underpinning. It is hoped that the thesis provides the underpinning from which this task may proceed in a more methodical way.

NOTES

1. Autonomy in the sense of acting on self-legislated principles also features prominently in Mill's thinking. But according to Mill, not forcibly preventing agents from acting on their voluntary choices is the best way to ensure that human beings flourish.
2. Reference to the self in terms of fundamental and peripheral desires and beliefs is neutral between reductionist and non-reductionist accounts of personal identity. The metaphysical issue emerges, however, in Chapter Two, as a complication in the discussion of subsequent alienation of the right to non-interference.
3. This is not to suggest that moral thinking as such necessarily involves a commitment to the universalizability principle, as Hare would have it. Taylor (1982) and MacIntyre (1982), for example, reject any such conception of moral thinking as too narrow. The present point is that it is part of our moral thinking about rights that we should make no distinctions without relevant differences.
4. J. J. Thomson, "Self-Defense and Rights." University of Kansas, Lindley Lecture, 1976, p.10. As cited in A. Gewirth (1984), p.91 n.1.
5. Brandt defines fully informed desires (which he calls rational desires) as those which survive maximal criticism and correction by facts and logic (1979, p.10).
6. Lectures on Well-Being, Trinity Term, Oxford, 1984.
7. Bricker (1980) gives an account of maximizing well-being (which he calls prudence) with respect to an agent's life as a whole, and derivatively to the individual acts that make up that life, in terms of a comparison between a totting up of (timeless) preferences satisfied in an agent's actual life, and a totting up of (timeless) preferences satisfied in other lives he could have lived if he had acted differently. Besides the problems of comparing different possible worlds, the idea of totting up in this context is notoriously difficult to make coherent. Griffin has argued, more plausibly, that for most purposes we do not need additive scales in order to maximize well-being. Instead, an ordinal scale of prudential value will suffice. (Lectures on Measurement in Morality, Hilary Term, Oxford, 1985.)
8. Murphy (1974) points out that some states in the U.S. do not distinguish between committability and incompetence by statute, and even where they are distinguished by statute,

the two judgements are often confused in practice.

9. If a subject is subsequently incapable of having an overriding desire that the act of paternalistic interference in question should or should not have occurred, perhaps because he lacks the necessary concepts to possess this sophisticated sort of desire, then for him the test of subsequent approval^e does not arise in the same way. We should not conclude, however, that in such cases the protection or promotion of the subject's well-being is sufficient in itself to justify paternalistic intervention. For that could lead to the unacceptable conclusion that it is permissible to give animals, severely retarded persons, the senile, and other permanent incompetents, mind-altering drugs if it could be successfully argued that their well-being would be thus enhanced. If we value autonomy, it cannot follow from the incapacity of subjects to alienate their right to non-interference, that their autonomy may be gratuitously overridden.
10. Mill's critics have often pointed out that there are no purely self-regarding actions. Nearly everything we do affects others in some way. This frequently heard contention does not detract, however, from the use of the notion of self-regardingness in the distinction between competence to decide what is in one's own best interests and competence to make one's own self-regarding decisions. Self-regarding actions are understood here simply as those in which the effects on others are morally either insignificant or irrelevant. Self-regarding considerations may be morally overridden from an all-things-considered point of view by the principle of preventing harm to others, but that is a different matter.
11. The internal/external distinction may be slightly misleading. With the exception of physical force, external conditions do not compel a subject unless he has certain internal desires which he cannot help acting from and which are closely connected to those external conditions.
12. It is also sometimes suggested, although the point is controversial, that certain desires by their very nature are intrinsically irrational. Murphy (1974) suggests that, for example, a masochistic desire for intense pain is intrinsically irrational (p.473, n.15).
13. Other reasons include:
 - (a) the protection of patients.
 - (b) the avoidance of fraud and duress.
 - (c) the encouragement of self-scrutiny by medical professionals.
 - (d) the promotion of rational decisions.

- (e) the involvement of the public in promoting autonomy as a general social value and in controlling biomedical research. (Beauchamp and Childress, 1979, p.63).
14. Quoted from mimeographed copy. Page numbers of published version not available.
 15. Feinberg says: "To whatever extent there is compulsion, misinformation, excitement or impetuousness, clouded judgement (as e.g. from alcohol), or immature or defective faculties of reasoning, to that extent the choice falls short of perfect voluntariness." (1971, pp. 110 - 111).
 16. Purportedly, in at least one place (Victoria, Australia) in which, prior to seatbelt legislation, it was argued that this legislation would have the effect of reducing medical costs, the opposite was found to occur after it was introduced. The unexpected increase in medical costs was attributed to a lower fatality rate among accident victims, and hence an increase in the number of persons requiring treatment.

BIBLIOGRAPHY

- Arneson, R. J. "Mill Versus Paternalism." Ethics 90 (1980): 470 - 489.
- Bayles, M. D. "A Concept of Coercion." In J. R. Pennock and J. W. Chapman (eds.) Coercion (Nomos 14) Chicago: Atherton, 1972, pp. 16 - 29.
- Bayles, M. D. "Criminal Paternalism." In J. R. Pennock and J. W. Chapman (eds.) The Limits of Law (Nomos 15) Chicago: Atherton, 1974, pp. 174 - 188.
- Beauchamp, T. L. "Paternalism and Biobehavioral Control." Monist 60 (1977): 62 - 80.
- Beauchamp, T. L. and Childress, J. F. Principles of Biomedical Ethics. Oxford: University Press, 1979.
- Benn, S.I. "Freedom, Autonomy and the Concept of a Person." Proceedings of the Aristotelian Society 76 (1976): 109 - 130.
- Benn, S. I. and Weinstein, W. L. "Being Free to Act and Being a Free Man." Mind 80 (1971): 194 - 211.
- Bergstroin L. "Utilitarianism and Alternative Actions." Nous 5 (1971): 237 - 252.
- Berlin, I. "Two Concepts of Liberty." In Four Essays on Liberty. Oxford: University Press, 1969, pp. 118 - 172.
- Blackburn, S. "Moral Realism." In J. Casey (ed.) Morality and Moral Reasoning. London: Methuen, 1971, pp. 101 - 124.
- Blackburn, S. "Truth, Realism and the Regulation of Theory." Midwest Studies in Philosophy 5 (1980): 353 - 371.
- Blackburn, S. "Rule Following and Moral Realism." In S. Holtzman and C. Leich (eds.) Wittgenstein: To Follow a Rule. London: Routledge and Kegan Paul, 1981.
- Bok, S. Lying. Sussex: Harvester, 1978.
- Bok, S. Secrets. Oxford: University Press, 1984.
- Brandon, E. P. "Rationality and Paternalism." Philosophy 57 (1982): 533 - 536.
- Brandt, R. B. A Theory of the Right and the Good. Oxford: Clarendon Press, 1979.

- Bricker, P. "Prudence." Journal of Philosophy 77 (1980): 381 - 401.
- Buchanan, A. "Medical Paternalism." Philosophy and Public Affairs 7 (1978): 370 - 390.
- Carter, R. "Justifying Paternalism." Canadian Journal of Philosophy 7 (1977): 133 - 145.
- Childress, J. F. "Who Should Decide? Paternalism in Health Care." Oxford: University Press, 1982.
- Cooper, D. E. "The Free Man." In A. P. Griffiths (ed.) Of Liberty. Cambridge: University Press, 1983, pp. 131 - 145.
- Culver, C. M. and Gert, B. Philosophy in Medicine. Oxford: University Press, 1982.
- D'Agostino, F. "Mill, Paternalism and Psychiatry." Australasian Journal of Philosophy 60 (1982): 319 - 330.
- Dearden, R. F. "Autonomy and Education." In R. F. Dearden, P. H. Hirst, and R. S. Peters (eds.) Education and Reason. (Part 3 of Education and the Development of Reason) London: Routledge and Kegan Paul, 1972, pp. 58 - 75.
- Dearden, R. F. "Autonomy as an Educational Ideal I." In S. C. Brown (ed.) Philosophers Discuss Education. London: Macmillan, 1975, pp. 3 - 18.
- Dennett, D. C. Elbow Room: The Varieties of Free Will Worth Wanting. Oxford: Clarendon Press, 1984.
- Devlin, P. "The Enforcement of Morals." Maccabaeen Lecture in Jurisprudence of the British Academy. Oxford: University Press, 1959.
- Devlin, P. The Enforcement of Morals. Oxford: University Press, 1965.
- Downie, R. S. and Telfer, E. Respect for Persons. London: Allen and Unwin, 1969.
- Dworkin, G. "Acting Freely." Nous 4 (1970): 367 - 383.
- Dworkin, G. "Paternalism." In R. A. Wasserstrom (ed.) Morality and the Law. Belmont, California: Wadsworth, 1971, pp. 107 - 126.
- Dworkin, G. "The Concept of Autonomy." Science and Ethics 12/13 (1981): 203 - 213.

- Dworkin, G. "Is More Choice Better Than Less?" Midwest Studies in Philosophy 7 (1982): 47 - 62.
- Dworkin, R. Taking Rights Seriously. London: Duckworth, 1977.
- Dworkin, R. "Is There a Right to Pornography?" Oxford Journal of Legal Studies 1 (1981a): 177 - 212.
- Dworkin, R. "What is Equality? Part I: Equality of Welfare." Philosophy and Public Affairs 10 (1981b): 185 - 246.
- Dworkin, R. "Rights as Trumps." In Jeremy Waldron (ed.) Theories of Rights. Oxford: University Press, 1984, pp. 151 - 167.
- Edwards, R. B. "Mental Health as Rational Autonomy." Journal of Medicine and Philosophy 6 (1981): 309 - 322.
- Elster, J. Ulysses and the Sirens. Cambridge: University Press, 1979.
- Elster, J. "Sour Grapes - Utilitarianism and the Genesis of Wants." In A. Sen and B. Williams (eds.) Utilitarianism and Beyond. Cambridge: University Press, 1983.
- Feinberg, J. "What is So Special About Mental Illness?" Doing and Deserving. Princeton: University Press, 1970, pp. 272 - 292.
- Feinberg, J. "Legal Paternalism." Canadian Journal of Philosophy 1 (1971): pp. 105 - 124.
- Feinberg, J. "The Idea of a Free Man." In J. F. Doyle (ed.) Educational Judgments. London: Routledge and Kegan Paul, 1973, pp. 143 - 172.
- Feinberg, J. Social Philosophy. New Jersey: Prentice-Hall, 1973.
- Feinberg, J. "Behaviour Control: Freedom and Behaviour Control." Encyclopaedia of Bio-Ethics 1 (1979): 93 - 100.
- Feinberg, J. "The Child's Right to an Open Future." In H. La Follette, (ed.) Children's Rights. Littlefield and Adams, 1980a.
- Feinberg, J. "The Interest in Liberty on the Scales." Rights, Justice and the Bounds of Liberty. Princeton: University Press, 1980b, pp. 30 - 44. First Published 1978.
- Feinberg, J. "The Rights of Animals and Unborn Generations." Rights Justice and the Bounds of Liberty. Princeton: University Press, 1980c, pp. 159 - 184. First Published 1974.

- Feinberg, J. Harm to Others. Oxford: University Press, 1984.
- Fingarette, H. Self-Deception. London: Routledge, 1969.
- Finnis, J. Natural Law and Natural Rights. Oxford: University Press, 1980.
- Finnis, J. Fundamentals of Ethics. Oxford: Clarendon Press, 1983.
- Fotion, N. "Paternalism." Ethics 89 (1979): 191 - 198.
- Frankfurt, H. G. "Alternate Possibilities and Moral Responsibility." Journal of Philosophy 66 (1969): 829 - 839.
- Frankfurt, H. G. "Freedom of the Will and the Concept of a Person." Journal of Philosophy 68 (1971): 5 - 20.
- Frankfurt, H. "Three Concepts of Free Action." Proceedings of the Aristotelian Society. Supplementary Volume 49 (1975): 113 - 125.
- Frey, R. G. "Rights, Interests, Desires and Beliefs." American Philosophical Quarterly 16 (1979): 233 - 239.
- Frey, R. G. Interests and Rights: The Case Against Animals. Oxford: Clarendon Press, 1980.
- Gert, B. "Coercion and Freedom." In J. R. Pennock and J.W. Chapman (eds.) Coercion (Nomos 14) Chicago: Atherton, 1972, pp. 30 - 48.
- Gert, B. and Culver, C. M. "Paternalistic Behavior." Philosophy and Public Affairs 6 (1976): 45 - 57.
- Gert, B. and Culver, C. M. "The Justification of Paternalism." Ethics 89 (1979): 199 - 210.
- Gert, B. and Culver, C. M. Philosophy in Medicine. Oxford: University Press, 1982.
- Gewirth, A. Reason and Morality. Chicago: University Press, 1978.
- Gewirth, A. "The Basis and Content of Human Rights." In J. R. Pennock and J. W. Chapman (eds.) Human Rights (Nomos 23) New York: University Press, 1981, pp. 119 - 147.
- Gewirth, A. "Are There Any Absolute Rights?" In J. Waldron (ed.) Theories of Rights. Oxford Readings in Philosophy Series. Oxford: University Press, 1984.
- Glass, M. "Not Going to Hell on One's Own." Philosophy 58 (1983): 471 - 480.

- Glover, J. "Assessing the Value of Saving Lives." In G. Vesey (ed.) Human Values. Hassocks, Sussex: Harvester Press, 1978, pp. 208 - 227.
- Glover, J. Causing Death and Saving Lives. Harmondsworth, Middlesex: Penguin, 1977.
- Glover, J. What Sort of People Should There Be? Harmondsworth, Middlesex: Penguin, 1984.
- Glover, J. "Self-Creation." Proceedings of the British Academy. Forthcoming, 1985.
- Gray, J. "John Stuart Mill on Liberty, Utility and Rights." In J. R. Pennock and J. W. Chapman (eds.) Human Rights (Nomos 23) New York: University Press, 1981, pp. 80 - 116.
- Gray, J. ^{and Pelczynski, Z. (eds.)} Conceptions of Liberty in Political Philosophy. London: Athlone Press, 1984.
- Gray, J. Mill on Liberty: A Defence. London: Routledge and Kegan Paul, 1983.
- Griffin, J. "Are There Incommensurable Values?" Philosophy and Public Affairs 7 (1977): 39 - 59.
- Griffin, J. "On Life's Being Valuable." Dialectics and Humanism 2 (1981): 51 - 62.
- Griffin, J. "Modern Utilitarianism." Revue Internationale de Philosophie 36 (1982): 331 - 375.
- Griffiths, A. P. (ed.) Of Liberty. Cambridge: University Press, 1983.
- Gutmann, A. "Children, Paternalism and Education: A Liberal Argument." Philosophy and Public Affairs 9 (1980): 338 - 358.
- Haksar, V. Equality, Liberty and Perfectionism. Oxford: Clarendon Press, 1979.
- Hamlyn, D. W. "Self-Knowledge." In T. Mischel (ed.) The Self. Oxford: Basil Blackwell, pp. 170 - 202.
- Hampshire, S. The Freedom of the Individual. London: Chatto and Windus, 1965.
- Hare, R. M. Freedom and Reason. Oxford: Clarendon, 1963.
- Hare, R. M. "Autonomy as an Educational Ideal." In S. C. Brown (ed.) Philosophers Discuss Education. Proceedings of the Royal Institute of Philosophy Conference, London, 1973, pp. 36 - 42.

- Hare, R. M. "What is Wrong With Slavery?" Philosophy and Public Affairs 8 (1979): 103 - 121.
- Hare, R. M. Moral Thinking. Oxford: Clarendon Press, 1981.
- Harman, G. The Nature of Morality. New York: Oxford University Press, 1977.
- Hart, H. L. A. "Immorality and Treason." The Listener. July 30, 1959, pp. 162 - 163.
- Hart, H. L. A. Law Liberty and Morality. Oxford: University Press, 1963.
- Hart, H. L. A. "Between Rights and Utility." In A. Ryan (ed.) The Idea of Freedom. Oxford: University Press, 1979, pp. 77 - 98.
- Held, V. "Coercion and Coercive Offers." In J. R. Pennock and J. W. Chapman (eds.) Coercion (Nomos 14) Chicago: Atherton, 1972, pp. 49 - 62.
- Hirst, P. H. "Liberal Education and the Nature of Knowledge." In R. F. Dearden, P. H. Hirst, and R. S. Peters (eds.) Education and Reason (Part 3 of Education and the Development of Reason) London: Routledge and Kegan Paul, 1972, pp. 1 - 24.
- Hirst, P. H. Knowledge and the Curriculum. London: Routledge and Kegan Paul, 1974.
- Hodson, J. "The Principle of Paternalism." American Philosophical Quarterly 14 (1977): 61 - 70.
- Hodson, J. "Mill, Paternalism, and Slavery." Analysis 41 (1981): 60 - 62.
- Hohfeld, W. N. Fundamental Legal Conceptions as Applied in Judicial Reasoning. Yale: University Press, 1919.
- Houlgate, L. D. The Child and the State: A Normative Theory of Juvenile Rights. Baltimore: John Hopkins University Press, 1980.
- Husak, D. N. "On the Rights of Non-Persons." Canadian Journal of Philosophy 10 (1980): 607 - 622.
- Husak, D. N. "Paternalism and Autonomy." Philosophy and Public Affairs 10 (1981): 27 - 46.
- Kohlberg, L. Stages in the Development of Moral Thought and Action. New York: Holt, Rinehart, and Winston, 1969.
- Kronman, A. T. "Paternalism and the Law of Contracts." Yale Law Journal 92 (1983): 763 - 789.

- Kuhn, T.S. The Structure of Scientific Revolutions. London: University of Chicago Press, 1970. First published 1962.
- Lee, S. "On the Justification of Paternalism." Social Theory and Practice 7 (1981): 193 - 203.
- Lively, J. "Paternalism." In A. P. Griffiths (ed.) Of Liberty. Cambridge: University Press, 1983, pp. 147 - 165.
- Locke, J. John Locke: Two Treatises of Government. P. Laslett (ed.) Cambridge, 1960. Originally published 1689.
- MacIntyre, A. After Virtue. London: Duckworth, 1981.
- Mackie, J. L. Ethics: Inventing Right and Wrong. Penguin: London, 1977.
- Mackie, J. "Can There Be a Right-Based Moral Theory." Midwest Studies in Philosophy 3 (1978): 350 - 359.
- McCloskey, H. J. "Rights." Philosophical Quarterly 15 (1965): 115 - 127.
- McCloskey, H. J. "Moral Rights and Animals." Inquiry 22 (1979): 23 - 54.
- McDowell, J. "Are Moral Requirements Hypothetical Imperatives?" Proceedings of the Aristotelian Society. Supplementary Volume 52 (1978): 13 - 29.
- McDowell, J. "Virtue and Reason." Monist 62 (1979): 331 - 350.
- McDowell, J. "Non-Cognitivism and Rule-Following." In S. Holtzman and C. Leich (eds.) Wittgenstein: To Follow a Rule. London: Routledge and Kegan Paul, 1981, pp. 141-162.
- Mill, J. S. On Liberty. C. V. Shields (ed.) New York: Bobbs-Merrill, 1956. First published 1859.
- Mill, J. S. Principles of Political Economy. In Collected Works 3 J. M. Robson (ed.) Toronto: University of Toronto Press, 1965.
- Minogue, K. R. "Freedom as a skill." In A. P. Griffiths (ed.) Of Liberty. Cambridge: University Press, 1983, pp. 197 - 215.
- Mischel, T. (ed.) The Self: Psychological and Philosophical Issues. Oxford: Basil Blackwell, 1977.
- Murphy, J. G. "Incompetence and Paternalism." Archiv fur Rechts und Sozialphilosophie 60 (1974): 465 - 486.

- Nagel, T. "The Limits of Objectivity." In S. McMurrin (ed.) Tanner Lectures on Human Values 1. Salt Lake City: University of Utah Press, 1980, pp. 79 - 118.
- Narreson, J. "Human Rights: Which, If Any, Are There?" In J. R. Pennock and J. W. Chapman (eds.) Human Rights (Nomos 23) New York: University Press, 1981, pp. 175 - 198.
- Neely, W. "Freedom and Desire." Philosophical Review 83 (1974): 32 - 54.
- Nozick, R. "Coercion." In S. Morgenbesser et al. (eds.) Philosophy, Science and Method. New York: St. Martin's, 1969, pp. 440 - 472.
- Nozick, R. Anarchy, State and Utopia. Oxford: Basil Blackwell, 1974.
- Nozick, R. Philosophical Explanations. Oxford: University Press, 1981.
- O'Neill, O. and Ruddick, W. (eds.) Having Children: Philosophical and Legal Reflections on Parenthood. Oxford: University Press, 1979.
- Parfit, D. "Personal Identity." Philosophical Review 80 (1971): 3 - 27.
- Parfit, D. Reasons and Persons. Oxford: Clarendon Press, 1984.
- Pateman, C. "Women and Consent." Political Theory 8 (1980): 149 - 168.
- Peters, R.S. "Freedom and the Development of the Free Man." In J. F. Doyle (ed.) Educational Judgements. London: Routledge and Kegan Paul, 1973, pp. 119 - 142.
- Rawls, J. A Theory of Justice. Cambridge: University Press, 1971.
- Rawls, J. "Rational and Full Autonomy." Journal of Philosophy 77 (1980): 515 - 535.
- Regan, D. H. "Justifications for Paternalism." In J. R. Pennock and E. Chapman (eds.) The Limits of Law (Nomos 15) Chicago: Atherton, 1974, pp. 189 - 209.
- Richards, D. A. "Rights and Autonomy." Ethics 92 (1981): 3 - 20.
- Rorty, R. Philosophy and the Mirror of Nature. Oxford: Basil Blackwell, 1980.

- Ryan, A. "Property, Liberty and On Liberty." In A. P. Griffiths (ed.) Of Liberty. Cambridge: University Press, 1983.
- Ryan, A. (ed.) The Idea of Freedom. Oxford: University Press, 1979.
- Sartre, J. P. Being and Nothingness. New York: Washington Square Press, 1966. First published 1943.
- Scarre, G. "Children and Paternalism." Philosophy 55 (1980): 117 - 124.
- Schoeman, F. "Responsibility and the Problem of Induced Desires." Philosophical Studies 34 (1978): 293 - 301.
- Schrag, F. "The Child in the Moral Order." Philosophy 52 (1977): 167 - 178.
- Scruton, R. "Freedom and Custom." In A. P. Griffiths (ed.) Of Liberty. Cambridge: University Press, 1983.
- Shope, R. K. "Rawls, Brandt, and the Definition of Rational Desires." Canadian Journal of Philosophy 8 (1978): 329 - 340.
- Singer, M. Generalization in Ethics. London: Eyre and Spottiswoods, 1963.
- Singer, P. Practical Ethics. Cambridge: University Press, 1979.
- Skinner, B. F. Beyond Freedom and Dignity. London: Jonathan Cape, 1972.
- Soble, A. "Paternalism, Liberal Theory and Suicide." Canadian Journal of Philosophy 12 (1982): 335 - 352.
- Sprigge, T. L. S. "Metaphysics, Physicalism and Animal Rights." Inquiry 22 (1979): 101 - 143.
- Strawson, P. "Freedom and Resentment." Freedom and Resentment and Other Essays. London: Methuen, 1974, pp. 1 - 25.
- Sumner, L. W. Abortion and Moral Theory. Princeton, New Jersey: University Press, 1981.
- Swinburne, R. and Shoemaker, S. Personal Identity. Oxford: Basil Blackwell, 1984.
- Taylor, C. "What is Human Agency?" In Mischel (1977) pp. 103 - 138.

- Taylor, C. "Responsibility for Self." In G. Watson (ed.) Free Will. Oxford Readings in Philosophy Series. Oxford: University Press, 1982, pp. 111 - 126.
- Taylor, C. "The Diversity of Goods." In A. Sen and B. Williams (eds.) Utilitarianism and Beyond. Cambridge: University Press, 1982, pp. 129 - 144.
- Taylor, C. "What's Wrong with Negative Liberty?" In A. Ryan (ed.) The Idea of Freedom. Oxford: University Press, 1979, pp. 175 - 194.
- Telfer, E. "Autonomy as an Educational Ideal II." In S. C. Brown (ed.) Philosophers Discuss Education. London: Macmillan, 1975, pp. 19 - 35.
- Ten, C. L. Mill On Liberty. New York: Oxford University Press, 1980.
- Thalberg, I. "Hierarchical Analyses of Unfree Action." Canadian Journal of Philosophy 8 (1978): 211 - 226.
- Thomson, J. "Rights and Compensation." Nous 14 (1980): 3 - 15.
- Tooley, M. "Abortion and Infanticide." Philosophy and Public Affairs 2 (1972): 37 - 65. Also "Correspondence" in Philosophy and Public Affairs 2 (1973): 419 - 432.
- Toulmin, S. E. "Self-Knowledge and Knowledge of the 'Self'." In Mischel (1977) pp. 291 - 317.
- Van de Veer, D. "Paternalism and Subsequent Consent." Canadian Journal of Philosophy 9 (1979): 631 - 642.
- Van de Veer, D. "Autonomy-Respecting Paternalism." Social Theory and Practice 6 (1980): 187 - 207.
- Vesey, Godfrey (ed.) Human Values. Hassocks, Sussex: Harvester Press, 1978.
- Waldron, J. (ed.) Theories of Rights. Oxford Readings in Philosophy Series. Oxford: University Press, 1984.
- Warnock, M. Existentialism. Oxford: University Press, 1970.
- Warnock, M. Imagination. London: Faber and Faber, 1976.
- Warnock, M. "Educating the Imagination." In Vesey (1978), pp. 44 - 60.
- Wasserstrom, R. A. (ed.) Morality and the Law. Belmont, California: Wadsworth, 1971.

- Watson, G. "Free Agency." Journal of Philosophy 72 (1975): 205 - 220.
- Watson, G. (ed.) Free Will. - Oxford Readings in Philosophy Series. Oxford: University Press, 1982.
- Winters, B. "Willing to Believe." Journal of Philosophy 76 (1979): 243 - 255.
- Young, R. "Compatibilism and Conditioning." Nous 13 (1979): 361 - 378.
- Young, R. "Autonomy and the 'Inner Self'". American Philosophical Quarterly 17 (1980): 35 - 43.
- Young, R. "Autonomy and Socialization." Mind (1980b) 89: 565 - 576.
- Young, R. "The Value of Autonomy." Philosophical Quarterly 32 (1982): 35 - 44.